Technical Report 1108

Specifications for an Operational Two-Tiered Classification System for the Army
Volume I: Report

Joseph Zeidner, Cecil Johnson, Yefim Vladimirsky, and Susan Weldon
The George Washington University

August 2000

United States Army Research Institute
for the Behavioral and Social Sciences

DTIC QUALITY INSPECTED 3

20010110 066

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

EDGAR M. JOHNSON
Director

NOTICES

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE August 2000	3. REPORT TYPE AND DATES COVERED Final Report: October 1996 - December 1997

4. TITLE AND SUBTITLE Specifications for an Operational Two-Tiered Classification System for the Army, Volume 1: Report	5. FUNDING NUMBERS C DASW01-96-M-2713 20363007A792 WU 1222

6. AUTHOR(S) Joseph Zeidner, Cecil Johnson, Yefim Vladimirsky, Susan Weldon (The George Washington University)	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The George Washington University Department of Administrative Sciences Washington, DC 20052	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences Attn: TAPC-ARI-RS 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SPONSORING / MONITORING AGENCY REPORT NUMBER Technical Report 1108

11. SUPPLEMENTARY NOTES
Technical Report consists of two volumes. Volume II, Appendices, is ARI Research Note 2000-12.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The broad objective of the present study is to design an improved two-tiered classification system and to compare its classification efficiency to the current operational aptitude area (AA) system. The total data set includes about 260,000 recruits serving in 170 different entry-level MOS during 1987-1989. The set includes all available ASVAB/Skill Qualification Test (SQT) data for MOS with adequate sample sizes collected by ARI during this time frame.

The proposed system to be evaluated in this study would use an invisible or black-box first tier in which separate assignment variables (AVs) are computed for up to 150 job families. The first tier AVs are to be used in assigning recruits to entry-level MOS. The second tier is used in recruiting, counseling and administration. The proposed system to be evaluated in the visible second tier uses up to 17 families. It is proposed that the aptitude area scores of the visible system be recorded on each soldier's personnel record.

The principal finding of the present study is that the unbiased overall mean predicted performance (MPP) of the 150 job family structure is .195 compared to the MPP for the existing operational system of .023, a gain of more than eight fold. The unbiased overall MPP for the 17 job families is .146 . The 17 family system is obtained by shredding the existing AA families within the boundaries of the operational classification families to maximize the Horst index.

Findings continue to support an early differential assignment theory (DAT) principle that maximum MPP is obtainable by using AVs for all jobs having adequate or stable validity data. The results clearly demonstrate that considerable classification efficiency is potentially obtainable from the existing ASVAB if it is used in accordance with DAT principles.

14. SUBJECT TERMS military personnel classification, Armed Services Vocational Aptitude Battery (ASVAB), classification efficiency, differential assignment theory, Army AA composites	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18 298-102

USAPPC V1.00

Specifications for an Operational Two-Tiered Classification System for the Army
Volume I: Report

Joseph Zeidner, Cecil Johnson, Yefim Vladimirsky,
and Susan Weldon
The George Washington University

Selection and Assignment Research Unit
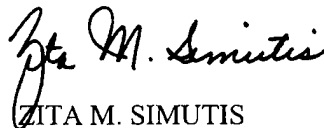**Michael G. Rumsey, Chief**

FOREWORD

The Army's enlisted selection and classification system has essentially been in place since 1949. It consists of a simple selection process using the Armed Forces Qualification Test to accept or reject potential recruits for military service, followed by a classification and assignment process using nine aptitude area (AA) composites. The current operational AA system of unit weighted, four-test composites and corresponding job families evolved in the Army, as with similar AA systems in the other military services, after many decades of research focusing almost entirely on enhancing predictive validity. The content of the Armed Services Vocational Aptitude Battery and of its composites have been primarily selected to maximize predictive validity rather than on improving classification efficiency in a multiple-job, optimal assignment context.

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has been conducting research over a number of years into ways of improving classification in the Army. The research reported here proposes a new set of classification-efficient composites. Use of these composites introduces a true performance metric into the classification process, and when combined with optimizing job-person matching procedures (such as the Enlisted Personnel Allocation System, EPAS) will result in substantial soldier performance gains. The changes recommended in the classification system are so profound and potential benefits so great that policy makers and personnel managers need to be made fully aware of the proposed two-tiered system.

The major recommendation made by the authors is to replace the current classification system with a two-tiered system. The two-tiered system has more than eight times the classification efficiency of the current AA system, with a mean predicted performance of .195 versus .023! The first tier consists of an invisible classification processor for comparing new composite scores within and across individuals and making classification recommendations for initial assignment of recruits. The first tier job family structure calls for 150 least-squares-estimates composites that are used as assignment variables. The scores themselves are not revealed to the recruits, but rather, a set of recommendations of possible job assignments are shown to them. In the visible second tier, a set of 17 composites are provided for recording onto each recruit's personnel record and for use when cut scores are required or for career counseling.

The development of the second tier composites form the basis for ARI's recommendation to replace the existing composites with 17 new AA composites and corresponding job families in December 2001 (when the numerical operations and coding speed subtests are dropped from the ASVAB under changes mandated by Office of the Secretary of Defense). The 150 predicted performance equations are integral to ARI's classification optimization research, exemplified by EPAS now undergoing field testing.

These research results and applications have been briefed to the Director of Military Personnel Management, the Deputy Chief of Staff for Personnel, and their staffs over the 1997 - 2000 period.

ZITA M. SIMUTIS
Technical Director

v

Acknowledgment

Executive Summary

Requirement

This report is one of a series of U.S. Army Research Institute-sponsored efforts by the authors concerning improving classification efficiency of the Armed Services Vocational Aptitude Battery through the use of least squares estimates of the criterion and through the restructuring of job families, making the families more homogeneous and more numerous.

In previous research, the authors demonstrated the classification efficiency (CE) of a number of alternative job family structures. A new two-tiered classification system, to replace the single-tiered operational aptitude area system, was proposed. The large gain in CE resulting from the use of a 66 job family structure led the investigators to conclude that a two-tiered system was highly feasible and that it was the most promising alternative being evaluated.

However, the least squares estimate (LSE) weights used in computing assignment variable (AVs) composites for some of the 66 job families were not based on large enough sample sizes to justify operational use. Also, it was clear that a new effort intended to study a proposed two-tiered system should include all available data in forming refined and more numerous job clusters to better represent the Army's total Military Occupational Specialty (MOS) structure.

The proposed system to be evaluated in this research would use an invisible or black-box first tier in which separate AVs are computed for up to 150 job families. The first tier AVs are to be used in assigning recruits to entry-level MOS. The second tier is used in recruiting, counseling and administration. The proposed system to be evaluated in the visible second tier uses up to 17 families. It is proposed that the aptitude area scores of the visible system be recorded on each soldier's personnel record.

The broad objective of the present research, then, is to define an improved two-tiered classification system and to compare its CE to the current operational aptitude area system. The total data set includes about 260,000 recruits serving in 170 different entry-level jobs during 1987-1989. The set includes all available ASVAB/Skill Qualification Test (SQT) data for MOS with adequate sample sizes collected by ARI during this time frame.

Procedures

An essential element of the research paradigm employed by the authors in a series of related investigations and the present one is the use of mean predicted performance (MPP) rather than the validity of predictor composites as the figure of merit. The measurement of MPP is generally achieved through the simulation of a two-stage selection and classification procedure. The first stage is a simple selection process in which a single selection variable is used to accept or reject candidates into an organization. The second stage is a classification and assignment process involving the matching of candidates to multiple jobs. The selection variable used in the first stage is usually a test composite (i.e., Armed Forces Qualification Test) believed to be a good predictor of performance in all jobs in the organization for which selection is being accomplished. The assignment variables correspond to the jobs or job families to which assignment of the selected individuals are made during the second stage. In the present research, however, only the Army input sample, rather than the total youth sample, is used in simulating the classification process. The simulation experiments conducted in the present research use empirical rather than synthetic scores.

Benefits obtained from optimally assigned recruits to MOS or job families are estimated under a number of conditions. Three independent samples are used to obtain: (1) regression weights for defining the assignment variables (AVs) that are maximized by the optimal assignment algorithm (Sample A); (2) regression weights for defining the evaluation weights (EVs) used to compute predicted performance scores (Sample B); and (3) test scores which are weighted to form both AVs and EVs which are in turn used for the simulations and determination of MPP after optimal assignment to jobs (Sample C). The MPP scores obtained from each simulation, separately for each experimental condition, are aggregated across independent cross-samples to measure system benefits.

Test intercorrelations and validities computed in both Samples A and B are corrected to provide estimates of the Army input population, and the simulations, in Sample C, use score vectors randomly selected from the same Army input population. Thus, regression weights and MPPs provided by the simulation experiments relate to the Army input population rather than the youth population.

The data used in this research are corrected for criterion unreliability and restriction in range, separately by MOS. Since restriction in range, when the Army input is designated as the population, is attributable only to the operational classification and assignment process, no correction is made for restriction due to the selection process. Separate LSE assignment composites are computed for a number of job families ranging in size from 25 to 150 for possible use in the first tier, and for 10, 13 and 17 job families for possible

use in the second tier. The CE of each family size is compared, in terms of MPP, to one another and to the current 10 operational AAs (including the small GT family).

Findings

The principal finding of the present research is that the unbiased overall MPP of the 150 job family structure is .195 compared to the MPP for the existing operational system of .023, a gain of more than eight times. The unbiased overall MPP for the 17 job families is .146 compared to the MPPs for the 10 and 13 families of .123 and .138 respectively. The 17 family system is obtained by shredding the existing AA families within the boundaries of the operational classification families to maximize H_d under this constraint. H_d, as defined by Horst, is an index equal to the average squared difference between each pair of criterion variables. This method uses the "constrained H_d" algorithm.

Findings continue to support an early differential assignment theory (DAT) principle that maximum MPP is obtainable by using AVs for all jobs having adequate or stable validity data. The results clearly demonstrate that considerable classification efficiency is potentially obtainable from the existing ASVAB if it is used in accordance with DAT principles.

Unbiased estimates of LSE equations in this and earlier investigations show clear improvements over the operational unit-weighted composites when sampling error is controlled through the use of cross-samples. Also, increasing the number of assignment composites and associated job families adds to potential CE. However, there is a much smaller increase in MPP as the number of job families are increased greatly (from 25 to 150 families), compared to larger increases in MPP as the number of families are increased moderately (from 10 to 25 families).

The research shows that the independent contribution to total MPP of classification is 56% compared to 44% for selection. Thus, the total process produces more than twice as much gain in MPP as the gain from selection alone. The existing AA composites, while having at best moderate value, have unacceptable negative MPPs for a number of individual job families. It is important to note that even increments as small as .1 in MPP measured in statistical standard scores translate into significant and practical estimates of monetary gain.

Operational Implications

A major recommendation made, based on both the findings of the present and earlier research, is to replace the current operational classification system with a two-tiered system. The first tier is used to propose initial job assignments for presentation to recruits by recruiting counselors. The computation and use of predicted performance to provide optimal assignments is transparent, occurring within the black-box. The output of the black-box is a computer-generated list of "best" job suggestions for presentation to the recruit. This output is based on predicted performance, minimum scores and other factors such as job quotas, quality distributions and policy constraints. The set of 17 family scores are used as cut scores for administration other than for initial assignment. The scores are also used for counseling and are recorded on each enlistee's personnel record. The aptitude area composites proposed for the second tier of the new system are subsets of operational families that are familiar to those now in current use and have scoring scales with the same meaning and values as found in the operational AAs.

The report also makes a number of research recommendations including the need for maintaining the proposed system with the passage of time; improving the race and gender fairness of the ASVAB and its composites; and improving the CE of the ASVAB by developing and using more classification-efficient tests.

Specifications for an Operational Two-Tiered Classification System for the Army

Table of Contents: Volume 1

List of Figures: Volume 1

Specifications for an Operational Two-Tiered Classification System for the Army

Table of Contents: Volume 2

[Note to the reader: Volume 2 is comprised of all the appendices to which references are made; it is published separately as an ARI Research Note of the same name.]

Introduction

Need for the Research

In May 1995, the results of earlier research on alternative job family structures were reported by the authors (Johnson, Zeidner & Vladimirsky, in preparation). A new two-tiered classification system was proposed. Initially, the research included an evaluation of a 66 job family structure only as a means of determining the extent of MPP increase beyond the use of a 25 job family structure then being considered for a revised single-tiered system. The relatively high increase in classification efficiency obtained from the use of 66 job families led the investigators to conclude that a two-tiered system was highly feasible and that it was the most promising alternative being evaluated.

It was recognized, however, that the least squares estimate (LSE) weights based on the full set of predictors, for some of the 66 job families were not based on large enough sample sizes to justify operational use. Additionally, it was clear that new research, intended to specify a proposed two-tiered system, should include all available ASVAB/SQT data in forming refined and more numerous first-tier job cluster to better represent the Army's total MOS structure.

Some parameters of the proposed new two-tiered classification system to be evaluated are described below.

> An invisible or black-box first tier is provided in which separate least squares estimate (LSE) assignment variables (AVs), based on the full set of ASVAB tests, are computed for 25 to 150 or more job families. The AVs are used in assigning recruits to entry-level MOS. A visible, revised second tier of 10 to 17 families, is provided for recruiting, counseling and administration. The aptitude area composite scores of the visible system are intended to be recorded on each soldier's individual record. The new research would determine the exact number of families in each tier.

> The remaining 80 or more jobs not included in the research would be linked to one of the families in the first tier, forming an expanded job family cluster. Each such job family cluster is centered on either a single MOS or a "core" of several MOS that has an adequate enough sample size to provide a stable LSE composite. Using the same linkage of 80-plus jobs to single or core MOS, the extended job families of the second tier, or visible system, are formed.

Research Objectives

The broad goal of the present research research is to define an improved two-tiered enlisted classification system and to compare its classification efficiency (CE) to the current operational aptitude area system.

Specific objectives are:

1. To restructure the job family system, creating both the first and second tiers.

2. To provide an unbiased assessment of CE, in terms of MPP, for AVs and corresponding job families that incorporate a number of system design principles based on differential aptitude theory. System characteristics defined and measured include number of job families, method for clustering MOS into job families, and type of test composite used to make assignment to job families.

3. To develop and measure constrained solutions to the optimal allocation algorithm that result in eliminating MPP values for individual job families falling below the statistical standard score MPP means for the total set of job families.

4. To provide a procedure for using operational ASVAB scores directly in AV composites by use of transformation weights. The corrected scores would retain the amount of CE demonstrated in the simulation experiment for the first-tier and be compatible with the current system in the second or visible tier.

5. To determine the effect on MPP of using enlistee samples that vary in job experience from one to three years during the first tour of duty. If no significant differences were found, it would permit combining enlistee samples.

To accomplish these goals, a number of research parameters are defined, including: establishing a "black-box" first tier in which separate LSE assignment composites are computed for a number of job families ranging in size from 25 to 150 formed from 170 MOS; establishing 10, 13, and 17 job families for possible use in the second or visible tier; linking all the remaining 80 or so jobs not used in the analysis to one of the job families in the first tier, thus creating expanded job family clusters around each "core" family; and using the same linkages of the 80 jobs in the first tier for expanding the second tier, thus consistently assigning all MOS to both first and second tiers.

Prior Investigations

The present research is one of a series of research efforts aimed at improving the selection and classification systems of the ASVAB based on differential aptitude theory (DAT) principles. Brogden (1951, 1959) and Horst (1954) were early proponents of job specific predictor composites in the multiple job context, a key principle of DAT. More recent proponents include Johnson and Zeidner (1991, 1995); Johnson, Zeidner and Leaman (1992); Scholarios, Johnson and Zeidner (1994); Statman (1992); Zeidner and Johnson (1991a, 1991b, 1991c, 1994); and Zeidner, Scholarios and Johnson (1997).

As described by Zeidner and Johnson (1991a), DAT is consistent with findings of dominance of a measure of general mental ability, g, in providing composites with high validity. It contends, however, that studies focused only on assessing the credibility of specific aptitude theory in terms of predicted validity are inadequate for explaining the benefits of optimally weighted predictor composites for classification. It thus avoids justification or rejection of tailored predictor composites on the basis of predictive validity alone.

An essential element of the research paradigm developed to illustrate the benefits predicted by DAT proponents is the use of MPP rather than the validity of predictor composites. The measurement of MPP is achieved through the simulation of a two-stage selection and classification procedure. The first stage is a simple selection process in which a single selection variable is used to accept or reject candidates into an organization. The second stage is a classification and assignment process involving the matching of candidates to multiple jobs. The selection variable used in the first stage is usually a test composite believed to be a good predictor of performance in all jobs in the organization for which selection is being accomplished. The assignment variables correspond to the jobs or job families to which assignment of the selected individuals are made during the second stage.

Zeidner, Johnson and Vladimirsky (1998) established the adequacy of SQT measures of job knowledge to serve as surrogates for core technical proficiency (CTP) measures of hands-on job proficiency in developing classification procedures for ASVAB assignment variables. The SQT would be considered an adequate substitute for CTP if it could be shown that the same developmental decisions are reached or that equivalent findings or outcomes are obtained using either criterion. Decisions to be made include the selection of tests for best assignment composites and the determination of weights for these tests. The overall correlation between the two criterion measures is found to be .46, while the correlation between the predicted performance scores of the two criteria is .92. These findings suggest that the two criteria measure different things, but when the criteria are measured in the joint predictor-criterion or valid space, the clear indication is

3

that either criterion could serve as a surrogate for the other in making classification decisions concerning ASVAB.

In considering the tests selected for the best 5-test composites, either criterion would select nearly the same tests for each job family test composite and provides comparable multiple validities.

The mean predicted performance (MPP) is .223 using CTP and .239 using SQT. Also, the stability of MPPs in cross samples of different data sets is higher for SQT than for CTP. The pattern of test validities across ASVAB tests considered separately for the 15 MOS were also quite similar. The mean corrected validity for CTP is found to be .76 and for SQT the validity is found to be .83

A principal component analysis of CTP and SQT in the joint predictor-criterion space yields six rotated oblique factors that showed good simple structure. Most of the 15 pairs of MOS had their highest coefficients (loadings) on the same factor for both criteria.

The overall conclusion, then, is that either criterion could serve as a surrogate for the other in conducting classification research on ASVAB. Key decisions made using either job proficiency criterion were very similar and outcomes were judged to be equivalent.

Johnson, Zeidner and Vladimirsky (in preparation) provide evidence that the proposed new FLS test composites have at least as much fairness for black and female recruits as do the existing aptitude area composites. Fairness is traditionally defined as the absence of underpredictions for the minority groups in which discrimination potentially exists. Another major objective of the fairness research is the development of a new methodology for making decisions based on trade-offs between classification efficiency and fairness to minorities in the design and use of a new system that optimally assigns first-term recruits to jobs.

In evaluating test composite prediction error scores resulting from operational assignment to MOS and job families, a distinct pattern of underpredictions was found for blacks and females. In testing for statistical significance, the means of prediction errors for blacks were not found to have statistically significant differences from zero for the set of MOS at the .05 level. In testing the mean differences for females, these differences from zero for the set of MOS were found to be significant at the .05 level. More importantly, the prediction error differences for blacks and females were too small to have practical significance. For blacks, the overall mean prediction error was -.025 of a standard deviation, or 0.5 in Army aptitude area standard score units. Aptitude areas have a mean of 100 and an SD of 20. For females, the mean prediction error was -.086, or about 1.7 points in AA standard score units.

4

The findings for minorities are consistent with research findings in the civilian employment and military settings concerning regression-line differences in intercept values. Such differences, in the same direction as found in this research, appear to be a relatively common phenomenon. In instances of comparisons between groups, the use of regression equations computed on the total group gives advantage to minority group members, i.e., minority groups are overpredicted. In the fairness research, using the total sample, we find underpredictions for minorities. Because of the large proportion of minorities in the total sample, only small changes in regression weights were made. Comparisons of the existing AA composites with the proposed new LSE composites resulted in much smaller prediction errors for the new composites for all groups.

When prediction errors of optimized and constrained optimized assignments in the classification context were compared for the existing 9 family AA test composites, the new composites resulted in much smaller prediction errors. Again, there was a consistent pattern of underpredictions for minorities, but these small differences are of little practical consequence.

In the comparison of classification efficiency for two types of 9 family composites, the new LSE composites were found to have an overall mean predicted performance (MPP) three times greater than the existing AA composites. Further, the existing AA composites were found to have unacceptable negative MPPs for three of the nine job families compared to no negative family MPPs for the new 9 job family LSE composites. The percentage of blacks assigned to combat jobs for both AA and LSE composites was found to be comparable to the percentages assigned by the operational system.

The overall conclusion, then, is that the proposed new LSE composites are fair to minorities while providing substantial improvements in classification efficiency. LSE composites assign blacks to combat jobs proportionately and without resorting to the use of racial quotas. The proposed new composites produce outcomes far superior to the existing AA composites on all the significant indices including: classification efficiency, positive MPPs for all job families, and size of prediction errors.

Johnson, Zeidner and Vladimirsky (in preparation) confirm that the restructuring of job families in the Army's classification and assignment system has potential promise for the improvement of the Army's personnel classification system — an improvement second only to that obtainable from the substitution of full least squares estimates of the criterion for the present operational unit weighted test composites. Findings show that the potential utility obtainable from optimal assignment of recruits to job families is greatly increased as the number of job families is increased. The investigations provide strong technical

support for recommending a change in the current operational system which incorporates an increased number of restructured job families.

The research also showed that while the empirical classification-efficient clustering algorithm showed a substantial superiority to judgment based clustering when only 9 families are to be utilized, no superiority was in evidence as the number of job families reached 25. It would appear that for systems with more than a dozen job families, one can rely on clustering by judgment that considers the operational classification family and career management fields (CMF) membership, and to a lesser extent, other considerations. Similarly, families derived from the empirical clustering algorithm and then adjusted by "judgment" produced values of MPP that are only slightly smaller than the empirically optimized families.

Other more basic research findings include the discovery that inflation in biased designs increases as the number of job families increases, making it inappropriate to use a biased design to estimate the effect of number of job families on MPP. The triple cross analysis design, previously developed by the authors and utilized in this research to produce the findings identified as being unbiased, should continue to be utilized in simulation experiments conducted to determine classification efficiency.

Methodology

Simulation of Operational Classification Systems

The Classification Concept

Initial classification from the Army input sample, without consideration of the selection process, is the sole concern of this research. Potential classification efficiency is estimated by simulation of a system in which the assignment of a recruit to a job family optimizes the sum of all recruits' AVs corresponding to the family to which each person is assigned. A linear programming (LP) algorithm is used to maximize this total sum of AVs as the objective function. This maximization of the objective function is accomplished under the constraint of meeting quotas for each assignment target, set proportionately to the accession numbers for the job families included in the analyses.

Optimal assignments to maximize the overall objective function can also be constrained to equal or exceed a prescribed objective function value in each job family. This additional "quality control" constraint was investigated in previous investigations directed at either defining feasible operational job families or evaluating the gender and racial fairness of test composites.

System Simulation

The Representation of Recruits by Entities

The entities utilized in a simulation can be obtained by either generating synthetic scores or by sampling from a real data set of individuals possessing scores for predictor variables relevant to the classification process being evaluated. While most of the authors' previous simulation investigations have used synthetic scores, primarily because the simulation of selection from a youth population requires such scores, this research relies entirely on empirical score vectors randomly assigned to the cross sample. The artificial persons, or entities, represented by these score vectors do not retain any additional characteristics, such as race, gender, previous history, or present assignments.

These entities are utilized only in the independent (cross) sample and are not corrected for attenuation or restriction in range. The truncation of the lower end of test distributions by AFQT selection and the censoring of the upper tail of this distribution by the reluctance of higher scoring youths to enlist is reflected in the score distributions, as are the variances and intercorrelations of test variables among these artificial persons. The individual scores of the score vectors of the same entities are weighted to obtain both

7

assignment variables (AVs) and evaluation variables (EVs) by using appropriate, independent weights. In unbiased designs, the sources of the EV weights differ from the sources of the AV weights, both sets of weights being drawn from mutually independent samples (here called samples A and B).

Because of the increased complexity in our research design, one that uses empirical rather than synthetic scores for creating test score vectors representing the youth population, and since we have sets of test scores representing the soldiers entering the Army that can be randomly assigned to the cross sample (here called sample C), we choose to use this entry population as our target population, for correcting sample A and sample B validities for restriction in range due to assignment. This contrasts with the synthetic score research designs where the youth population is the target population for restriction in range corrections of both predictor intercorrelations and validity coefficients. Since we are not generating synthetic scores, we do not have an automatic duplication of the shape (normal distribution) of the youth population.

In the research design for this research, using empirical scores, samples A and B are corrected only for restriction in range due to assignment effects--an effect due to the restriction in range impact of assignment to MOS from a common entry pool. We also correct the validities for unreliability of the criterion variables, prior to the restriction in range corrections. The test score vectors constituting entities in the cross samples, sample C, used for the experimental simulations are not corrected in any way.

Simulating the Classification Process

Since only conclusions relating to potential classification efficiency are presented in the conclusions of this research, there is no need to simulate all actual operational classification procedures. We do not simulate procedures that reflect: minimum cut scores, travel costs, recruit preferences, or several other factors that can assume considerable importance in an operational system. Only optimal assignment to jobs or job families and filling job quotas (numbers assigned to jobs) are simulated in this series of investigations. Constraints on the optimization process are imposed to represent quotas and to improve quality distribution to families. The authors believe implementation of a drastically changed set of job families and corresponding AVs of a redesigned classification system can be rationally proposed because of the significant gain in potential classification efficiency that can be obtained from such a redesign.

Assignments to job families are determined in one or more simulation subsamples used as experimental cross samples. In the most highly biased design, the same data set may be used for the analysis, evaluation, and simulation subsamples. However, in the completely unbiased design--the one most often

8

used in this research–a triple cross-validation design is utilized. The triple cross-validation design calls for the use of two independent subsamples for the computation of regression weights for AVs and for EVs.

A third independent subsample is used as the source of the score vectors (entities) to which these weights are applied to obtain new and totally unbiased AVs and EVs. This third sample is used for the actual conduct of the simulation and evaluation processes. The analysis or evaluation sample data used to compute regression weights for either AVs or EVs are corrected first for attenuation and then for a restriction in range effect caused by the classification and assignment process. The entities in the third (simulation) sample are not corrected in any way.

In this research design, one which uses empirical test scores instead of synthetic test scores in our cross samples, as noted earlier, we do not correct intercorrelation and validity matrices obtained in the analysis and evaluation samples for restriction in range due to selection effects, but do correct for restriction in range due to assignment effects from a common entry pool. Also as noted, validities are corrected for unreliability of the criterion variables. The regression weights applied to entities in the simulation sample to obtain both AVs and EVs are computed from these corrected matrices.

Alternative Simulation Concepts

Alternative approaches to the identification of job families and corresponding test composites were evaluated by the authors using a simulation paradigm in the context of several possible operational systems. Each simulation investigation must stipulate a proposed type of operational system to guide the construction of the assignment procedures. A number of such system types are described below and is followed by a discussion of the relative merits of these alternative types.

(1) Two Stage System: Selection Followed by Selection (Selection-Selection or Multiple Hurdle Model).

The personnel system that preceded EPAS may be characterized as one in which initial selection into the organization is followed by the use of a second stage selection process. In this system, assignment of new recruits to MOS occurred on the basis of soldier preferences, needs of the service (including the availability of training seats), whether each recruit exceeds a minimum cut score on the test composites reflecting job requirements for each family of MOS, and to varying extent on such cost factors as travel distances. In actuality, a recruit's configuration of ASVAB test scores has little effect on the assignment decisions in a system so heavily dependent on stated preferences and minimum cut scores. Higher and more divergent cut

9

scores would need to be used for the ASVAB test scores in this type of system to have non-trivial impact on the assignment decision.

(2)   Two Stage System: Selection Followed By Classification.

A one stage selection process, or a two stage selection process as described above, may be followed by a classification stage where an effort is made to assign each recruit to the MOS in which the recruit is expected to do best and/or a particular recruits performance can be expected to be of maximum value to the organization. The classification stage can be accomplished using a single measure of general intelligence (g) multiplied by the differing validities of g against performance in the various MOS. Alternatively, test composites can be used that maximally predict performance in the corresponding job families. When only g is used to provide optimal assignment, the classification process is referred to as hierarchical classification; when multiple test composites are used in the optimal assignment process the classification process is referred to as allocation.

The classification process can in turn be separated into either two tiers or two echelons. An example of the use of two echelons was provided in simulation research closely related to the present one, the authors optimally assigned entities to the nine operational job families (the first echelon) followed by optimal assignment to the MOS within each job family (the second echelon). These two echelons can be accomplished sequentially and differ from the two tiered classification approach in that certain constraints are applied during the assignments to MOS. The assignment of an individual to his/her optimal MOS is restricted to those MOS in the job family to whyich he/she was assigned in the first echelon. The two classification echelons could be accomplished in the first tier of the two tiered operational system; the first echelon and the second tier are conseptually the same. Thus, the second echelon could be embedded in the second tier with ACSS utilized for both job families and MOS–but at a considerable cost in MPP. The sizable increase in MPP resulting in such a two-echelon, single tiered, classification system–as compared to the results obtained from the use of only a single echelon assignment to the nine families–is a reflection on the poor homogeneity among the MOS within the operational job families.

(3)   Two Tiered Classification System: Selection Followed By Classification and again by Classification.

The classification process that follows a first stage selection decision may be divided into two operational tiers. The first of these tiers is a black box (invisible or transparent) procedure in which initial

assignment recommendations relate to a relatively large number of kernels. These kernels are mostly comprised of single MOS with validity data based on large enough samples to permit computing stable regression weights to form best weighted test composites. These composites are used as assignment variables in the first tier. A smaller number of MOS with only small validity data samples; have these samples combined to permit computing stable LSEs, with the centroids of the MOS clusters designated as kernels.

The second tier, as used in the present investigation, is based on a smaller number of job families obtained by clustering first tier kernels (i.e. MOS or clusters of MOS) so as to maximize a measure of classification efficiency, H_d. Initially, sets of 9, 12 and 16 sets of kernels were considered in terms of CE and credibility, but a set of 17 kernels was finally selected. The test composite scores computed as LSEs of the criterion for second tier job families are intended to be recorded in the official record of each recruit and utilized in much the same way as in the second stage of the selection-selection system described above.

(4) A One Stage Selection-Classification System: Simultaneous Selection and Classification.

When it is feasible to make the selection and classification decisions on a single occasion, the multidimensional Screening (MDS) algorithm, described in Johnson and Zeidner (1991) and further demonstrated and evaluated by Whetzel (1991), makes it possible to achieve a higher MPP than when an optimal two stage selection-classification algorithm is utilized. However, the two-stage algorithm provides a number of advantages not obtainable when using a one-stage algorithm. The two-stage algorithm permits the use of distinct predictor variables for the selection and classification stages and, more importantly, permits the validation of predictors against distinct criterion variables for the two stages. The criterion variables most appropriate for developing predictors for a single stage may not be the most appropriate for either selection or classification stages. Because of these limitations, the MDS based operational alternative is not considered in the present investigation.

The Criterion

Evaluation of CE is conducted using predicted performance (i.e., the evaluation variable) based on the same set of predictor variables used to compute AVs. This approach follows Brogden's recommendation (1946; 1951; 1959) for the use of predicted performance as a substitute for unobtainable actual performance across the set of families to which optimal assignment is to be applied. In the unbiased designs of this

research, the AVs are independent of the evaluation variables, whether or not the AVs are also LSEs of the criterion; the regression weights for LSEs used as AVs are computed using an analysis sample while the regression weights for LSEs used as EVs are computed using evaluation sample data.

The substitutability of predicted performance for actual performance in the evaluation of both selection and classification efficiency of personnel systems–as well as the development and research relating to selection and classification systems–was suggested by both Horst (1954) and Brogden. Brogden (1955) provided a proof of this substitutability using several highly justifiable basic assumptions. Abbe (1968) established the robustness of Brogden's proof using a model sampling approach.

Criteria for the present research are SQTs for 170 MOS obtained during 1987-1989. The set of SQT scores in each of these MOS were standardized to have a mean of zero and a standard deviation of one within a single MOS. The figure of merit representing classification efficiency is the mean predicted performance, or, when expressed as a mean of statistical standard scores, MPP.

Statistical Tests of Significance

Most of the answers to the question of features that should be incorporated into an improved classification system can be found by determining which levels in the facets described below should be included in the system. The first step in this determination is the measurement of MPP provided at each level and the costs of incorporating this level in terms of other management goals. The investigators believe that statistical significance of the contribution of these three facets is already well established from the results from previous investigations. The goal in the present research is to explore differences in MPP across the most relevant levels within each of these facets.

The same sets of entities are utilized across all conditions in the experimental simulations, suggesting the use of a repeated measure factorial design or critical ratios in which the correlations of predicted performance across two conditions are included in the computation of the standard error. The investigators have chosen to compute critical ratios for those contrasts for which statistical significance is in doubt.

Experimental Conditions of Previous Investigations that became Key Design Features of a Proposed Operational System

Effect of Number of Job Families

Since MPP has been shown many times to increase as the number of job families increase, it is necessary to control this variable in order to compare the effectiveness of alternative system features. Conditions were simulated separately for 9, 16, 25, and 66 families in one prior investigation utilizing an unbiased design by the present authors. All except the largest of these sizes of families were also evaluated in Johnson, Zeidner, and Leaman (1992), but in a biased design. In each of these investigations it was found that MPP increases as the number of job families increases.

Clustering Methods

Several clustering methods were explored in the prior investigations on the effect system features on CE, including the following: (1) *a priori*, based on operational job families for CMF or initial classification, previous factor analyses bearing on the clustering of Army MOS, and/or expert judgment; (2) psychometric factors rotated to simple structure; (3) maximization of H_d (differential); and (4) the maximization of H_d, adjusted by judgment. A detailed description of the third clustering method, one which maximizes H_d, is provided in Johnson, Zeidner, and Leaman (1992).

Type of AV

The unit weighted operational ASVAB tests provide the baseline against which gains provided by assignment variables (AVs) computed as least square estimates (LSEs) of the criterion variable are compared.

Evaluating Simulation Outcomes

Subsamples of the Army Input Sample

The sample of recruits come from 170 selected MOS who had taken SQTs during 1987-1989. Their ASVAB test scores come from approximately the same period. Those enlistees who had taken the SQT only once during the four year period prior to 1991 are identified for a selected set of eleven MOS. The investigators did not know at the beginning of this research whether those who had taken the SQT only once were different in important ways from those who took the SQT more than once during the period of interest.

Entities were formed from the score vectors containing ASVAB test scores, with associated SQT scores, and MOS. These entities were used in the same way as synthetic scores to conduct simulation experiments in our earlier investigations. Then as now, the final output consists of MPPs obtained after assignment to job families. The total data set for the main investigation contained 260,000 cases. Two

13

random subsamples, stratified by the 170 MOS, each contain 120,000 cases, and a third random subsample contains 20,000 cases. The two equal sized subsamples are referred to as samples A and B respectively, and the smaller subsample as sample C. (See Figure 1.) The aggregate of samples A, B, and C is the designated population, the Army Input sample, and is referred to as P. The designated population–for designs in which synthetic scores are generated to create some or all of samples A, B, and C–is comparable to the total of samples A, B and C of this investigation.

Research Design

The generic simulation paradigm of this research can be described in terms of the following roles given to empirical samples of entities: (1) Sample A has the analysis role in which MOS are clustered into job families, weights for AVs are computed; in research with broader objectives the role of sample A would include the selection of tests for use in batteries or composites; (2) Sample B has the evaluation role in which weights are computed for LSEs, using all predictor information and SQT scores in each MOS or job family, to provide the best estimate of criterion scores; (3) Sample C has the simulation role in which entities are optimally assigned to job families using weights for AVs computed in an analysis sample and MPPs computed using weights for the evaluation variables (EVs) computed in the evaluation sample. See Figure 2.

It should be noted that predictor intercorrelations and validities are computed within MOS samples and then aggregated to provide similar information on job families. Only sample C utilizes individual entities. Entities consist of the predictor score vectors (without criterion scores or MOS identification). AVs and EVs are computed for each entity in order to assign entities to job families, and to evaluate results of the assignments in terms of MPP.

The alternative research designs range from being completely unbiased through varying degrees of bias. The amount of inflation in various types of biased designs (with respect to obtained MPPs) were

STEP ONE

(A)	(B)	(C)
EVALUATION SAMPLE N = 120,000	ANALYSIS SAMPLE N = 120,000	CROSS SAMPLE N = 20,000

COMPOSITE WEIGHTS (by job/family)

EVALUATION WEIGHTS (by job)

MPP RESULTS

STEP TWO

(B)	(A)	(C)
EVALUATION SAMPLE N = 120,000	ANALYSIS SAMPLE N = 120,000	CROSS SAMPLE N = 20,000

COMPOSITE WEIGHTS (by job/family)

EVALUATION WEIGHTS (by job)

MPP RESULTS

STEP THREE

(A) + (B) + (C)

SELECTION OF TESTS
AND WEIGHTS FOR
OPERATIONAL USE

N = 260,000
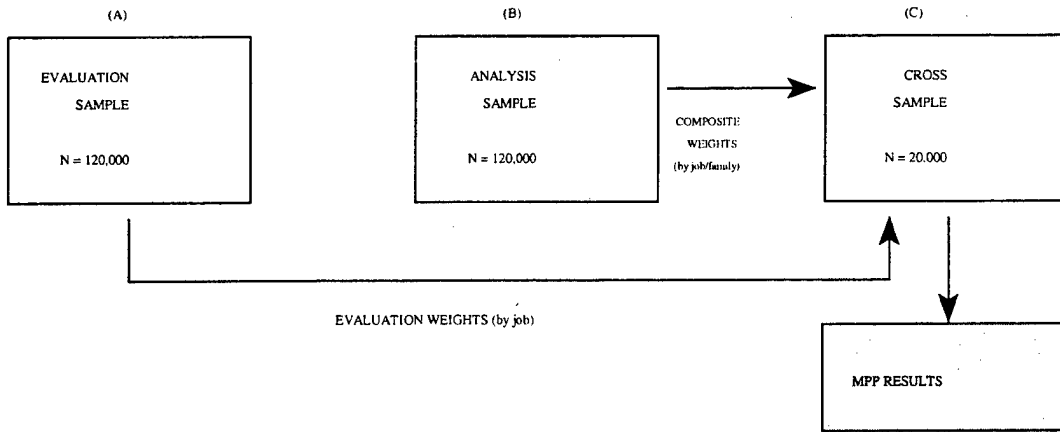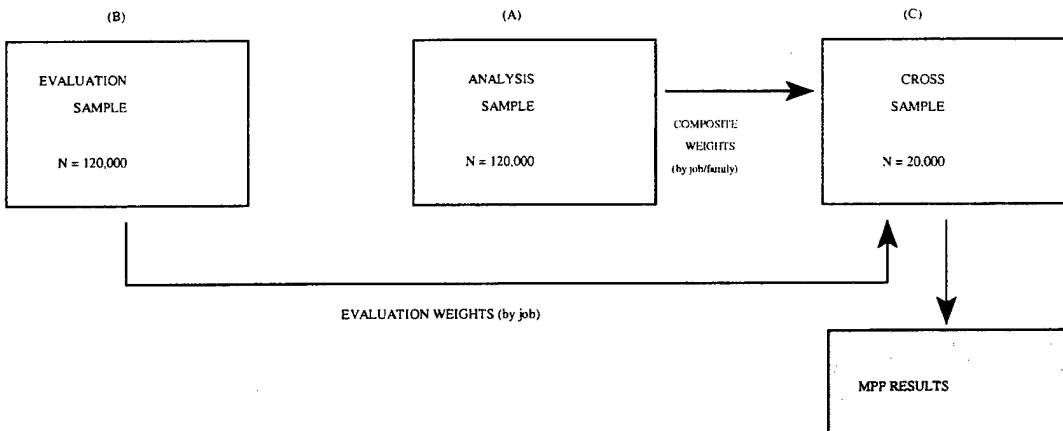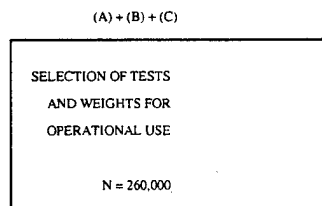
Figure 1 Triple Cross Analysis Design Sample Sizes

15

Total Empirical Sample

Evaluation Sample

Cross Validation Sample

Analysis Sample

Computation of MOS weights to be used for evaluation variables

Formation of random replication samples

Test selection and computation of family weights to be used for assignment variables

MPP RESULTS
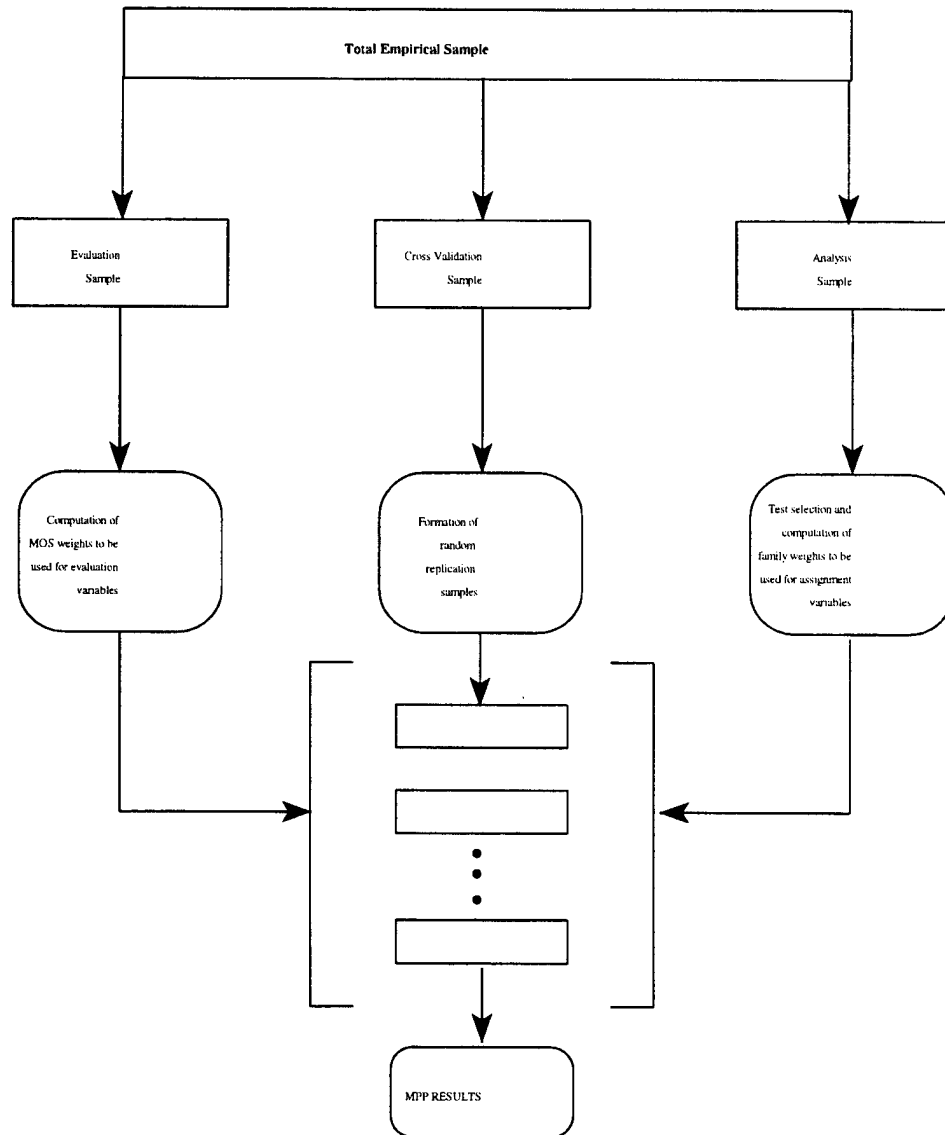
Figure 2 Generalized Research Design

investigated in an associated project. The simulation designs in this research can be described by indicating the sample which is used for each of the three roles described above (using the same order). Thus an A,B,C design is one in which sample A plays the analysis sample role, sample B plays the evaluation sample role, and C plays the evaluation sample role. Similarly, a P,P,P design (not used to compute MPP in this research) is one in which the same sample, P, plays all three roles.

Unbiased Estimates of MPP

Designs A,B,C and B,A,C provide MPP values that are not inflated by capitalization on sampling error. Additional stability of results can sometimes be obtained by averaging the results across A,B,C and B,A,C. This averaging can be readily accomplished when an *a priori* set of job families are in use and samples A and B provide only the regression weights for the AVs in the same families. However, when samples A and B are used to separately cluster jobs into job families, the results are essentially from two investigations and require a more sophisticated method for combining results.

Biased Estimates of MPP

Designs A,A,A or B,B,B or P,P,P use the same data source to compute AVs and EVs, providing correlated error to the computation of MPPs. Also, the more traditional "back-sample" inflation source is present because the same data is being used to compute regression weights and to make assignments to job families. Although the comparison of *a priori* composites with LSE composites could be misleading, there are comparisons where the increased stability provided by the larger Ns of the PPP design would over shadow the presence of shrinkage. Such biased estimates will be utilized in this research to provide the best possible weights for use in defining operational AVs.

Unbiased Design that Utilizes the Designated Population

The A,P,C design for using empirical scores is the one that most closely resembles the "unbiased" design previously used by the authors when synthetic scores are generated from the parameters of the designated population. Synthetic scores for analysis (A) and simulation (C) samples are generated, while EVs are computed from the parameters of the designated population. The empirical score A,P,C design is equivalent to the synthetic score design if P is assumed to represent the population.

17

MOS Kernels, Core MOS, and Job Families

Structuring of MOS into first tier job families is accomplished by first identifying single MOS kernels. Each MOS designated as a kernel has adequate validity data to permit the computation of reasonably stable LSEs for use as AVs for assignment purposes. The remainder of the approximately 250 Army MOS that are included in our total data sample were placed in an analysis set of MOS that included a few of the single MOS kernels that had the smaller sample sizes. The MOS in this analysis set were then clustered to maximize CE (i.e., Horst's H_d) until MOS groups with large enough samples with validity data resulted. The centroids of these MOS clusters were then added to the list of first tier kernels. Those MOS for which we either lacked data or had inadequate data to support a clustering process have been attached by judgment to one of the kernels. This provides first tier job families that include all Army MOS to which recruits may be initially assigned.

Conceptually, there are two kinds of kernels: single MOS with adequate validity data, and the centroids of clusters of MOS that individually would not possess sufficient validity data to provide stable LSEs. The LSEs computed against the criterion variable for each kernel are used to obtain validity coefficients in a cross sample. Also, the variances-covariances among these LSEs are computed in the sample in which the clustering to maximize CE is obtained. This variance-covariance matrix is the basic input matrix for the H_d clustering algorithm, the result of which will be trial tier-2 job families for evaluation in an unbiased simulation experiment.

In previous investigations by the authors, in which psychometric factors provided one option for identifying job families, the kernel MOS which load highly on a particular factor and much lower on all other factors were referred to as the core MOS for that factor. Similarly, two or more kernel MOS selected to represent a job family because it is judged that they form a relatively tight cluster in the center of a job family may be referred to as core MOS. Tentative core MOS, of either type, can be confirmed as core MOS if they have high correlations with the centroid of their own cluster and a distinctly lower correlation with the centroids of all other identified clusters.

As indicated above, all non-kernel MOS are attached to a specified kernel to form first tier families that include all initial input MOS. LSEs computed using only the validity data associated with the kernel would operationally be assumed to represent both the kernel and all of the attached MOS. However, in the research conducted in this investigation, no research use is made of the MOS attached to the kernels.

18

Classification Efficient Methods for Clustering MOS into Families for Tier 2 and for Clustering MOS with Small Sample Sizes

Basic Classification Efficient (CE) Clustering Algorithm

Johnson, Zeidner, and Leaman (1992) demonstrated a job clustering algorithm that provides an approximate maximization of classification efficiency measured in terms of MPP. Several modifications of this algorithm were tried out during this investigation.

The CE clustering algorithm used as the starting point in this research is described in greater detail and demonstrated in Johnson, Zeidner, and Leaman (1992). Leaman programmed this algorithm in FORTRAN. The modifications of this basic algorithm, as required in this investigation, were made to an equivalent program written in Gauss. The change in programming language was made to increase the ease with which matrix algebra expressions could be converted into computer code.

The CE clustering algorithm starts with the assumption that each MOS is a job family. CE (i.e. the H_d index) is maximized when the number of families is as large as possible. Operationally, this would occur when each MOS is treated as a family (i.e.,assignments are made separately to each MOS). Thus, H_d is at a maximum at the start of the CE clustering algorithm, and with each reduction in the number of families the CE is reduced. At each step two selected job families are combined (agglutinated) into a single job family. The pair of job families to be agglutinated at each step reflects the combination that makes a minimum reduction in H_d. The pair of families agglutinated during a step may be single MOS or one or both may be the centroids of MOS clusters that resulted from previous steps.

The modifications to this algorithm permit the investigator to stipulate sets of MOS that should not be agglutinated into the same family. Preprocessing can merge MOS into a family prior to use of the basic algorithm, regardless of whether the original or the modified program is utilized. These two capabilities facilitate imposing judgment on the empirical clustering process.

The clustering program can be used to test the effects of merging jobs in the preprocessing stage. Selecting several clusters of 2, 3 or 4 jobs that are logically linked and found in the same families in the output at just before and just after the desired number of families in the initial solution is reached, we can do the indicated merging in the preprocessing stage and rerun the algorithm so as to select the same number of job families.

The modified H_d job clustering algorithm also provides researchers with other capabilities that make this program into a more flexible job clustering tool. Researchers and managers are provided the capability of obtaining a job structure that avoids three problems: (1) having logically related jobs, jobs which policy considerations demand their being retained in the same family, ending up in different families; (2) having the algorithm agglutinate two or more clusters of jobs–that are highly acceptable with respect to logic and/or policy as distinct job families–into the same job families, while large number of jobs (MOS) still remain as distinct families (i.e., unassigned to an acceptable job family; and (3) having too many families at a possible stopping point where many unassigned jobs would eventually become members of already identified job families rather than becoming viable core job families, especially when the sample sizes on which the AVs for some of these "families" would be based are too small to provide stable LSEs for use as operational AVs.

In addition to a solution to these three problems, the usefulness of the present algorithm for creating operational viable job families required a capability to consider accession density of each MOS in the clustering algorithm. The previous FORTRAN program did provide for such weighting in the algorithm, but used weights that assumed equal accessions for every MOS. It was relatively simple to change the built in weights from equal weights to accession based weights for each MOS.

The MOS for constituting each central cluster, the core MOS, can be identified in a number of ways including: (1) policy or judgment of similarity among the MOS in the joint predictor-criterion space; (2) examination of factor analysis solutions to determine such similarities; and (3) examination of the complete sequence of clustering decisions as made by the clustering algorithm. The final output of the CE job clustering program (by itself) does not provide much information as to which MOS should be considered to be the core MOS. However, the clustering program's final output can be used to confirm that a set of MOS being considered for a pre-processing merge into a core family would also be placed in the same family by the clustering algorithm.

Classification Efficient (CE) Solutions Constrained to be Consistent with Operational Families

The above, unconstrained, CE algorithm has been modified so that all clusters are constrained to fall within the confines of the existing operational classification families, with no membership of MOS of one operational family presently placed in other families. This has the effect of dividing the present operational families into a larger number of smaller families using a method which maximizes the CE obtainable by adding one more family to the set, except for the constraint that all family members are from the same existing operational job family.

20

Judgment Based Clustering (*A Priori* Clusters)

Clustering MOS by judgment always makes use of knowledge of the membership of each MOS in CMFs and the existing classification families, as well as other research information that may be available. Two previous research investigations, in which the CE of sets of families identified using the H_d algorithm were compared with judgement based families, provided results which supported the concept that the difference in CE provided by judgement and empirical clustering decreases as larger numbers of families are utilized. No 9-family judgment clusters were identified in these previous investigations, although the operational classification families provided a 9-family *a priori* set. CMF and the existing classification family membership is frequently inconsistent and thus considerable room for judgment exists for the identification of *a priori* family systems containing 16 or more families. Such judgments were made primarily by one of the investigators who was familiar with several previous investigations that provided information on similarities among MOS. Included in those previous investigations was information regarding MOS that appeared to be more similar to MOS clusters in other job families than to those in the family they had been operationally placed.

Adjusted Empirical Clusters

The same kind of information used to create judgment clusters was also used to identify MOS in the empirical clusters that violated the concepts the investigators felt were appropriate for the creation of judgment clusters.

Model of the Operational System Proposed for Implementation

A number of the characteristics of the proposed operational system are highly relevant to the design of experimental simulations intended to evaluate alternative design features. The goal of the experimental simulations is to evaluate the effect on CE of changes in various system features. When other considerations are equal, a research simulation experiment which best replicates the design characteristics of the proposed operational system also best provides the most accurate evaluation of the proposed operational system. However, the imposition of possible, or even probable operational design details reflecting policy considerations that are inconsistent with obtaining the largest potential CE, are neither practical nor otherwise desirable for the investigators to include in experimental simulations. Also, the use of less optimal assignment algorithms to provide savings in the computing process were not considered desirable for inclusion in the experimental simulations. Thus, the estimates of CE provided by the reported simulations

21

generally pertain to the potential CE obtainable from an operational system which emphasizes maximization of predicted performance without separate consideration of preferences or travel costs, although one simulation experiment is designed to determine the loss in CE from use of a proposed method for providing a desirable quality distribution.

Traditional single-tiered personnel classification uses one set of job families and corresponding assignment variables (AV) to effect all classification procedures and vocational counseling. In the two-tiered operational system proposed and defined as one product of this research, all MOS to which initial classification can be made are designated as members of one of the job families of a set of 150 for tier one and a set of 17, linked to the first tier families, for tier two. During initial classification and tentative assignment to a MOS proposed as an operational system, an LP algorithm is used to maximize an objective function consisting of the sum of the first tier AVs corresponding to each MOS to which a soldier is assigned, subject to all constraints, including quotas, minimum cut scores, and probably quality constraints, being met. It is proposed that the values of minimum cut scores will be obtained for second tier AVs in essentially the same way as such cut scores are presently obtained for aptitude area scores. The designers of a two tiered operational system have the options of requiring, as a prerequisite to assignment, either: (1) having the first tier AV scores exceed the first tier cut scores that correspond to the second tier cut scores; or (2) having the second tier cut scores exceed a score that corresponds to the second tier cut score for the job family that includes the MOS for which assignment is being considered. Quality constraints, unlike cut scores, should be defined and enforced in terms of the first tier families and corresponding AVs.

Operationally, when a recruit is assigned to a first tier family, the assignment to a second tier family is automatic, since the second tier families are defined in terms of the constituent first tier families. Only the second tier AV scores are included in each soldier's record that is available to him/her, to counselors, and to his commanding officer. Counseling and self determination of eligibility for various Army schools and programs would be based on a soldiers scores on the second tier test composites, that is, on the 17 AVs plus a pass/fail code that indicates whether the minimum cut score for a specific MOS has been reached or exceeded.

A number of highly relevant characteristics of the proposed operational system have been or are being examined in this research using experimental simulations. These system characteristics include: (1) the nature of the test weights used to form AVs; (2) the means and standard deviations of AV scores after conversion to the scales used to make assignments; (3) the best kind and use of minimum cut scores

22

(Johnson, Zeidner, and Scholarios, in preparation). The proposed first tier system has the advantage over prior operational systems in that minimum cut scores would be applied in terms of AVs for 17 more homogeneous job families instead of 9 or 10 test composites corresponding to the existing heterogeneous aptitude area job families.

Our proposed operational system includes the provision of quality control across job families in terms of the AVs, instead of using AFQT as the measure of quality. Quality control is implemented in this proposed system being modeled in the simulation experiments by placing constraints on the minimum MPP permitted for each job family during the process in which the objective function (mean predicted performance over the total batch) is being maximized by an LP algorithm. Thus quality control is in terms of LSEs of performance in a job family, instead of general ability as measured by AFQT scores or the unit weighted composites utilized in the current operational system that are also primarily measures of g.

Applying the recommended quality control process to a system which uses the present, unit weighted, AAs as assignment variables causes the off the shelf LP algorithm in the General Algebraic Modeling System (GAMS) to send the message that there is no feasible solution. This situation apparently occurs because of the higher intercorrelations and the lower validities of the unit weighted AAs compared to the AVs consisting of LSEs. If much more severe quality control constraints were applied in an LP algorithm while using LSEs as the AVs, a problem would also arise and this message from GAMS would occur.

Findings of the Central Experiment: The Two-Tiered System

Sample Sizes

The first analysis undertaken was to obtain the means and standard deviations for each ASVAB test by MOS and the means of the SQTs by MOS after standardizing each MOS by year before combining years. This was done separately for the two samples, A and B, that were divided using a random number generator program. The test validities were first corrected for attenuation and then for restriction in range for the Army input population and separately for the youth population. These values are shown in Appendices A1 and A2 for the two samples. As mentioned in the previous section, there are about 120,000 enlistees in each of the two samples. An additional 20,000 enlistees comprise Sample C, the cross sample.

Classification Efficient (CE) Job Families

A previous investigation (Johnson, Zeidner & Vladimirsky, in preparation) suggested that the first tier of a two-tiered system should identify as many MOS as have adequate validity as distinct single MOS job families. There are 170 MOS with validity data in our data base that could have been initially considered, on the basis of sample size, in the process of creating job families. However, a number of these MOS with relatively smaller sizes were judged to be too small to stand as single MOS. The H_d empirical clustering algorithm was then employed to assist in identifying stable single MOS and multi-MOS job families. These small MOS were joined with other related small MOS to form eight separate combined MOS.

At the start of the H_d clustering algorithm, each single MOS is considered a job family. At each step two families are combined (agglutinated) into a single job family, with the pair of job families selected for agglutination producing a minimum reduction in H_d. Job families selected for initial consideration as the possible first tier are shown in Table 1. The combined MOS are designated as Zs in Table 1 and their constituent MOS are also listed. For example, job family 23 is labeled 24Z and it comprises four MOS (24C, 24G, 24N and 21L). The overall family structure comprises 146 single MOS (including five Z families) and 4 multi-MOS families (including two Z families). In the case of the example, 24Z, it is considered a single MOS family after smaller MOS were combined.

Table 1

The 150 Job Family First-Tier System

Family	N	MOS	Title
1	5000	11B	Infantryman
2	5000	11C	Indirect Fire Infantryman
3	5000	11H	Heavy Anti-Armor Weapons Infantryman
4	4593	11M	Fighting Vehicle Infantryman
5	5000	12B	Combat Engineer
6	1950	12C	Bridge Crewmember
7	603	12F	Engineering Tracked Vehicle Crewman
8	5000	13B	Cannon Crewmember
9	720	13C	Tacfire Operations Specialist
10	1919	13E	Cannon Fire Direction Specialist
11	4101	13F	Fire Support Specialist
12	776	13M	Multiple Launch Rocket Sys (MLRS) Crewmember
13	2724	13N	Lance Crewmember
14	592	13R	Fa Firefinder Radar Operator
15	683	14D	Hawk Missile Crewmember
16	703	16E	Hawk Fire Control Crewmember
17	1104	16P	Chaparral Crewmember
18	1996	16R	Vulcan Crewmember
19	2406	16S	Man Portable Air Defense System Crewmember
20	5000	19D	Cavalry Scout
21	4764	19E	M48-M60 Armor Crewman
22	5000	19K	M1 Abrams Armor Crewman
23	752	**24Z** 24C 24G 24N 21L	**Combined** Hawk Firing Section Mechanic Hawk Information Coordination Center Mechanic Chaparral System Mechanic Pershing Electronics Repairer
24	358	25S	Still Documentation Specialist
25	898	27E	TOW/Dragon Repairer

25

The 150 Job Family First-Tier System

Family	N	MOS	Title
26	852	29V	Strategic Microwave Systems Repairer
27	5000	31C	Single Channel Radio Operator
28	5000	31K	Combat Signaler
29	2778	31L	Wire Systems Installer
30	709	31N	Communications Systems/Circuit Controller
31	563	31P	Microwave Systems Operator-Maintainer
32	1394	31Q	Tactical Satellite/Microwave System Operator
33	5000	31R	Multichannel Transmission Systems Operator
34	498	31S	Satellite Communications System Operator
35	4278	31V	Unit Level Communications Maintainer
36	1021	35E	Radio and Communications Security Repairer
37	307	35H	TMDE Maintenance Support Specialist
38	1034	35J	Telecommunications Terminal Device Repairs
39	737	35N	Wire Systems Equipment Repairer
40	1201	36M	Switching Systems Operator
41	323	41C	Fire Control Instrument Repairer
42	1045	44B	Metal Worker
43	592	44E	Machinist
44	612	45B	Small Arms Repairer
45	565	45D	Self-Propelled FA Turret Mechanic
46	546	45E	M1 Abrams Tank Turret Mechanic
47	817	45K	Tank Turret Repairer
48	448	45L	Artillery Repairer
49	563	45N	M60A1/A3 Tank Turret Mechanic
50	509	45T	Bradley Fighting Vehicle Sys Turret Mech
51	498	**46Z**	**Combined**
		46Q	Journalist
		46R	Broadcast Journalist
52	2037	51B	Carpentry and Masonry Specialist

The 150 Job Family First-Tier System

Family	N	MOS	Title
53	532	51K	Plumber
54	327	51M	Firefighter
55	723	51R	Interior Electrician
56	344	51T	Technical Engineering Specialist
57	529	52C	Utility Equipment Repairer
58	5000	52D	Power Generator Equipment Repairer
59	1380	54B	Chemical Operations Specialist
60	2457	55B	Ammunitions Specialist
61	415	55D	Explosive Ordinance Disposal (EOD) Spec
62	791	57E	Laundry and Bath Specialist
63	3054	62B	Construction Equipment Repairer
64	1522	62E	Heavy Construction Equipment Operator
65	527	62F	Crane Operator
66	887	62J	General Construction Equipment Operator
67	5000	63B	Light-Wheel Vehicle Mechanic
68	1234	63D	Self-Propelled Field Artillery Sys Mech
69	1376	63E	M1 Abrams Tank System Mechanic
70	785	63G	Fuel and Electrical System Repairer
71	2396	63H	Track Vehicle Repairer
72	1302	63J	Quartermaster and Chemical Equip Repairer
73	750	63N	M60A1/A3 Tank System Mechanic
74	2506	63S	Heavy-Wheel Vehicle Mechanic
75	3378	63T	Bradley Fighting Vehicle Sys Mechanic
76	3062	63W	Wheel Vehicle Repairer
77	987	63Y	Track Vehicle Mechanic
78	1359	67N	Utility Helicopter Repairer
79	236	67R	AH-64 Attack Helicopter Repairer
80	1564	67T	Tactical Transport Helicopter Repairer

The 150 Job Family First-Tier System

Family	N	MOS	Title
81	1632	67U	Medium Helicopter Repairer
82	1751	67V	Observation/Scout Helicopter Repairer
83	1168	67Y	AH-1 Attack Helicopter Repairer
84	640	68B	Aircraft Powerplant Repairer
85	740	68D	Aircraft Powertrain Repairer
86	712	68F	Aircraft Electrician
87	904	68G	Aircraft Structural Repairer
88	1128	68J	Aircraft Armament/Missile Systems Repairer
89	388	68M	Aircraft Weapon Systems Repairer
90	900	68N	Avionic Mechanic

91	324	**68Z**
		68L
		68Q
		68R
			Combined
			    Avionic Communications Equipment Repairer
			    Avionic Nav & Flight Control Equipment Repairer
			    Avionic Special Equipment Repairer

92	1431	71D	Legal Specialist
93	1145	71G	Patient Administration Specialist
94	5000	71L	Administrative Specialist
95	972	71M	Chaplain Assistant
96	1651	72E	Tactical Telecommunications Center Op
97	1738	72G	Automatic Data Telecommunications Center Op
98	2246	73C	Finance Specialist
99	500	73D	Accounting Specialist
100	1184	74B	Information Systems Operator
101	4113	75B	Personnel Administration Specialist
102	2505	75C	Personnel Management Specialist
103	2714	75D	Personnel Records Specialist
104	1379	75E	Personnel Actions Specialist
105	624	75F	Personnel Information Sys Mgt Specialist
106	997	76J	Medical Supply Specialist

The 150 Job Family First-Tier System

Family	N	MOS	Title
107	2897	76P	Material Control and Accounting Specialist
108	5000	76V	Material Storage and Handling Specialist
109	541	76X	Subsistence Supply Specialist
110	5000	77F	Petroleum Supply Specialist
111	805	77W	Water Treatment Specialist
112	331	81L	Printing and Bindery Specialist
113	808	82C	Field Artillery Surveyor
114	1525	88H	Cargo Specialist
115	5000	88M	Motor Transport Operator
116	1954	88N	Traffic Management Coordinator
117	5000	91A	Medical Specialist
118	748	91D	Operating Room Specialist
119	1209	91E	Dental Specialist
120	474	91F	Psychiatric Specialist
121	309	91G	Behavioral Science Specialist
122	1478	91K	Medical Laboratory Specialist
123	513	91M	Hospital Food Service Specialist
124	695	91P	X-Ray Specialist
125	682	91Q	Pharmacy Specialist
126	558	91R	Veterinary Food Inspection Specialist
127	514	91S	Preventive Medicine Specialist
128	345	91T	Animal Care Specialist
129	641	**91Z** 91H 91J 91U 91Y	**Combined** Orthopedic Specialist Physical Therapy Specialist Ear, Nose and Throat Specialist Eye Specialist
130	5000	92A	Automated Logistical Specialist
131	5000	92G	Food Service Specialist

The 150 Job Family First-Tier System

Family	N	MOS	Title
132	298	92M	Mortuary Affairs Specialist
133	1009	92R	Parachute Rigger
134	5000	92Y	Unit Supply Specialist
135	626	93C	Air Traffic Control (ATC) Operator
136	1327	93P	Flight Operations Coordinator
137	5000	95B	Military Police
138	323	95C	Corrections Specialist
139	818	96B	Intelligence Analyst
140	361	96D	Imagery Analyst
141	792	96R	Ground Surveillance Systems Operator
142	429	97B	Counterintelligence Agent
143	562	98C	Signals Intelligence Analyst
144	1242	98G	EW Signal Intelligence Voice Interrogator
145	966	98H	Morse Interceptor
146	463	**98Z**	**Combined** (98D, 98J, 98K)
		98D	Emitter Locator/Identifier
		98J	Noncommunications Interceptor/Analyst
		98K	Non-Morse Interceptor/Analyst
147			
	215	55G	Nuclear Weapons Specialist
	303	93F	Field Artillery Meteorological Crewmember
148			
	548	**27Z**	**Combined**
		24K	Hawk Continuous Wave Radar Repairer
		24M	Vulcan System Mechanic
		27H	Hawk Firing Section Repairer
		27M	Multiple Launch Rocket System Repairer
		27N	Forward Area Alerting Radar (FAAR) Repairer
	433	**29Z**	**Combined**
		29F	Fixed Communications Security Equipment Repairer
		29M	Tactical Satellite Microwave Repairer
149			

The 150 Job Family First-Tier System

Family	N	MOS	Title
	451	25M	Graphics Documentation Specialist
	372	**25Z**	**Combined**
		25C	Cartographer
		25P	Visual Information/Audio Documentation Specialist
	372	97E	Interrogator
150			
	224	15E	Pershing Missile Crewmember
	171	16J	Defense Acquisition Radar Operator

Note: The unconstrained H_d empirical clustering algorithm was used to assist in defining the multi-MOS families

Job Families Modified by Judgment

Four types of job clustering are used in this research: (1) pure H_d algorithm (unconstrained and unmodified), with no judgment modification of resulting families; (2) constrained H_d based on shredding the operational AA families, also with no judgment modification of resulting families; (3) constrained H_d, with the resulting families moderately modified by judgment; and (4) constrained H_d, with the resulting families slightly modified by judgment.

As noted earlier, clustering MOS by judgment always makes use of knowledge of the membership of each MOS in career management fields (CMF) and in existing classification families. Based on earlier findings, a procedure was employed that imposed constraints on the H_d algorithm so that MOS could be agglutinated to form more acceptable (rational) families. The principal constraint initially imposed on H_d for two family sizes (17 and 13) was that all MOS included in a family would belong to the same operational family. Then the resulting families were moderately modified by judgment to improve the homogeneity of clustering. A third set of 17 job families (different from the first set of 17 families), identified by the constrained H_d clustering algorithms, was only slightly modified by judgment to remove the most obvious inconsistencies and to improve the homogeneity of job clusters. The operational classification of 10 families provides a third *a priori* set as a baseline for comparisons.

Table 2 shows the listing of jobs into the first 17 family set, where after the families were clustered by the constrained H_d they were moderately modified. An exception to the constraint of shredding with operational families is the inclusion of two MOS (Journalist and Broadcast Journalist) into a Science and Technology (ST) family rather than retaining their membership in the small GT job family. The listing of jobs in the operational 10-family set is shown in Appendix B of Volume 2. [Note to the reader: Volume 2 is comprised of all the appendices to which references are made; it is published separately as an ARI Research Note of the same name as this volume.]

Table 2

The 17 Family Second-Tier System

Cluster	MOS	Job Title
		Clerical Administration (CL) 1
1	71D	Legal Specialist
	71G	Patient Administration Specialist
	71L	Administrative Specialist
	71M	Chaplain Assistant
	73C	Finance Specialist
	73D	Accounting Specialist
	75B	Personnel Administration Specialist
	75C	Personnel Management Specialist
	75D	Personnel Records Specialist
	75E	Personnel Actions Specialist
	75F	Personnel Information Sys Mgt Specialist
	75H	Personnel Services Specialist
	76P	Material Control and Accounting Specialist
	88N	Traffic Management Coordinator
		Clerical Administration (CL) 2
2	76J	Medical Supply Specialist
	76V	Material Storage and Handling Specialist
	76X	Subsistence Supply Specialist
	77F	Petroleum Supply Specialist
	92A	Automated Logistical Specialist
	92Y	Unit Supply Specialist
		Combat (CO) 1
3	11B	Infantryman
	11C	Indirect Fire Infantryman
	11H	Heavy Anti-Armor Weapons Infantryman

Cluster	MOS	Job Title
	11M	Fighting Vehicle Infantryman
	18B	Special Forces Weapons Sergeant
	18X	Special Forces Candidate

Combat (CO) 2

4	12B	Combat Engineer
	12C	Bridge Crewmember
	12F	Engineering Tracked Vehicle Crewman
	18C	Special Forces Engineer Sergeant
	19D	Cavalry Scout
	19E	M48-M60 Armor Crewman
	19K	M1 Abrams Armor Crewman

Electronics (EL) 1

5	14E	Patriot Fire Control Enhanced Operator/Maintainer
	14L	AN/TSQ-73 Air Defense Command and Control Operator/Maintainer
	14M	Man Portable Air Defense System Crewmember (RC)
	18E	Special Forces Communications Sergeant (Special Operators Communications Spec)
	24Z	**Combined**
	24C	Hawk Firing Section Mechanic
	24G	Hawk Information Coordination Ctr Mech
	24N	Chaparral System Mechanic
	21L	Pershing Electronics Repairer
	25L	AN/TSQ-73 ADA Cmnd/Control System Operator/Repairer
	31L	Wire Systems Installer
	31R	Multichannel Transmission Systems Operator
	31V	Unit Level Communications Maintainer
	51R	Interior Electrician

The 17 Family Second-Tier System

Cluster	MOS	Job Title
	52G	Transmission and Distribution Specialist
	68M	Aircraft Weapon Systems Repairer
	68X	AH-64 Armament/Electrical Systems Repairer

Electronics (EL) 2

Cluster	MOS	Job Title
6	25V	Combat Documentation/Production Specialist
	31K	Combat Signaler
	31N	Communications Systems/Circuit Controller
	31P	Microwave Systems Operator-Maintainer
	31Q	Tactical Satellite/Microwave System Op
	31S	Satellite Communications System Operator
	31U	Signal Support Systems Specialist
	35Y	Integrated Family of Test Equipment (IFTE) Operator/Maintainer
	36M	Switching Systems Operator
	39Y	FA Tactical Fire Direction Systems Specialist
	55G	Nuclear Weapons Specialist
	74G	Telecommunications Computer Operator/Maintainer
	93F	Field Artillery Meteorological Crewmember
	96H	Air Intelligence Specialist
	96R	Ground Surveillance Systems Operator

Electronics (EL) 3

Cluster	MOS	Job Title
7	25R	Visual Information Equipment Operator/Maintainer (Audio/Visual Equip Repairer)
	27E	TOW/Dragon Repairer
	27F	Vulcan Repairer
	27G	Chaparral/Redeye Repairer
	27K	Hawk Fire Control/Continuous Wave Radar Repairer
	27T	Avenger System Repairer

The 17 Family Second-Tier System

Cluster	MOS	Job Title
	27X	Patriot System Repairer
	27Z	**Combined**
	24K	Hawk Continuous Wave Radar Repairer
	24M	Vulcan System Mechanic
	27H	Hawk Firing Section Repairer
	27M	Multiple Launch Rocket System Repairer
	27N	Forward Area Alerting Radar (FAAR) Repairer
	29H	Automatic Digital Message Switch Equipment Repairer
	29V	Strategic Microwave Systems Repairer
	29Z	**Combined**
	29F	Fixed Communications Security Equip Repairer
	29M	Tactical Satellite Microwave Repairer
	31F	Network Switching Sys Op/Maintainer (Mobile Subscriber Equip Network Sys Op)
	33R	Electronic Warfare/Intercept Aviation Systems Repairer
	35B	Land Combat Supply Systems Test Specialist
	35D	Air Traffic Control Equipment Repairer
	35E	Radio and Communications Security Repairer
	35H	TMDE Maintenance Support Specialist
	35J	Telecommunications Terminal Device Repairs
	35L	Avionic Communications Equipment Repairer
	35N	Wire Systems Equipment Repairer
	35Q	Avionic Flight Systems Repairer
	39B	Automatic Test Equipment Operator/Maintainer
	45G	Fire Control Repairer
	68J	Aircraft Armament/Missile Systems Repairer
	68N	Avionic Mechanic
	68Z	**Combined**
	68L	Avionic Communications Equipment Repairer

The 17 Family Second-Tier System

Cluster	MOS	Job Title
	68Q	Avionic Nav & Flight Control Equip Repairer
	68R	Avionic Special Equipment Repairer
	93D	ATC Systems, Subsystems and Equipment Repairer

Field Artillery (FA)

Cluster	MOS	Job Title
8	13B	Cannon Crewmember
	13C	Tacfire Operations Specialist
	13E	Cannon Fire Direction Specialist
	13F	Fire Support Specialist
	13P	MLRS/Lance Operations/Fire Direction Specialist

General Maintenance (GM) 1

Cluster	MOS	Job Title
9	41C	Fire Control Instrument Repairer
	44B	Metal Worker
	44E	Machinist
	45B	Small Arms Repairer
	45D	Self-Propelled FA Turret Mechanic
	45K	Tank Turret Repairer
	45L	Artillery Repairer
	45T	Bradley Fighting Vehicle Sys Turret Mech
	52C	Utility Equipment Repairer
	52D	Power Generator Equipment Repairer
	52F	Turbine Engine Driven Generator Repairer

General Maintenance (GM) 2

Cluster	MOS	Job Title
10	43M	Fabric Repair Specialist
	51B	Carpentry and Masonry Specialist
	51K	Plumber

The 17 Family Second-Tier System

Cluster	MOS	Job Title
	51M	Firefighter
	55B	Ammunitions Specialist
	55D	Explosive Ordinance Disposal (EOD) Spec
	57E	Laundry and Bath Specialist
	62E	Heavy Construction Equipment Operator
	62F	Crane Operator
	62G	Quarrying Specialist
	62H	Concrete and Asphalt Equipment Operator
	62J	General Construction Equipment Operator
	77W	Water Treatment Specialist
	88H	Cargo Specialist
	92M	Mortuary Affairs Specialist
	92R	Parachute Rigger

Mechanical Maintenance (MM) 1

Cluster	MOS	Job Title
11	24T	Patriot Operator and System Mechanic
	45E	M1 Abrams Tank Turret Mechanic
	45N	M60A1/A3 Tank Turret Mechanic
	62B	Construction Equipment Repairer
	63B	Light-Wheel Vehicle Mechanic
	63D	Self-Propelled Field Artillery Sys Mech
	63E	M1 Abrams Tank System Mechanic
	63G	Fuel and Electrical System Repairer
	63H	Track Vehicle Repairer
	63J	Quartermaster and Chemical Equip Repairer
	63N	M60A1/A3 Tank System Mechanic
	63S	Heavy-Wheel Vehicle Mechanic
	63T	Bradley Fighting Vehicle Sys Mechanic

38

Cluster	MOS	Job Title
	63W	Wheel Vehicle Repairer
	63Y	Track Vehicle Mechanic
	88K	Watercraft Operator
	88L	Watercraft Engineer
	88P	Railway Equipment Repairer (RC)
	88Q	Railway Car Repairer
	88R	Airbrake Repairer
	88S	Locomotive Electrician
	88T	Railway Section Repairer (RC)
	88U	Railway Operators Crewmember
	88V	Train Crewmember

Mechanical Maintenance (MM) 2

Cluster	MOS	Job Title
12	67G	Utility Airplane Repairer
	67H	Observation Airplane Repairer
	67N	Utility Helicopter Repairer
	67S	Helicopter Repairer
	67T	Tactical Transport Helicopter Repairer
	67U	Medium Helicopter Repairer
	67V	Observation/Scout Helicopter Repairer
	67X	Heavy Lift Helicopter Repairer
	67Y	AH-1 Attack Helicopter Repairer
	68B	Aircraft Powerplant Repairer
	68D	Aircraft Powertrain Repairer
	68F	Aircraft Electrician
	68G	Aircraft Structural Repairer

The 17 Family Second-Tier System

Cluster	MOS	Job Title

Operators & Food (OF)

Cluster	MOS	Job Title
13	13M	Multiple Launch Rocket Sys (MLRS) Crewmember
	13N	Lance Crewmember
	14D	Hawk Missile Crewmember
	14J	Early Warning Systems Operator **(F)**
	14R	Sight Forward Heavy Crewmember **(F)**
	14S	Avenger Crewmember
	15E	Pershing Missile Crewmember
	16D	Hawk Missile Crewmember
	16E	Hawk Fire Control Crewmember
	16H	Air Defense Artillery Operator/Intelligence Assistant
	16J	Defense Acquisition Radar Operator
	16P	Chaparral Crewmember
	16R	Vulcan Crewmember
	16S	Man Portable Air Defense System Crewmember
	16T	Patriot Missile Crewmember
	16X	Feeds 16B, 16D, 16E and 16T (CMF 16 Trainee)
	88M	Motor Transport Operator
	91M	Hospital Food Service Specialist
	92G	Food Service Specialist

Surveillance & Communication (SC)

Cluster	MOS	Job Title
14	13R	Fa Firefinder Radar Operator
	13T	Remotely Piloted Vehicle Crewmember
	31C	Single Channel Radio Operator
	72E	Tactical Telecommunications Center Op
	72G	Automatic Data Telecommunications Center Op
	74C	Telecommunications Operator/Maintainer

Cluster	MOS	Job Title

Skilled Technical (ST) 1

Cluster	MOS	Job Title
15	18D	Special Forces Medical Sergeant
	42C	Orthotic Specialist
	42E	Optical Laboratory Specialist
	77L	Petroleum Laboratory Specialist
	91A	Medical Specialist
	91B*	Medical NCO (called Medical Equipment Repairer, 91A, on CMF listing)
	91D	Operating Room Specialist
	91E	Dental Specialist
	91F	Psychiatric Specialist
	91G	Behavioral Science Specialist
	91K	Medical Laboratory Specialist
	91N	Cardiac Specialist
	91P	X-Ray Specialist
	91Q	Pharmacy Specialist
	91R	Veterinary Food Inspection Specialist
	91S	Preventive Medicine Specialist
	91T	Animal Care Specialist
	91V	Respiratory Specialist
	91X	Mental Health Specialist
	91Z	**Combined**
	91H	Orthopedic Specialist
	91J	Physical Therapy Specialist
	91U	Ear, Nose and Throat Specialist
	91Y	Eye Specialist

Cluster	MOS	Job Title
		Skilled Technical (ST) 2
16	25M	Graphics Documentation Specialist
	25S	Still Documentation Specialist
	25Z	**Combined**
	25C	Cartographer
	25P	Visual Info/Audio Documentation Specialist
	33T	Electronic Warfare/Intercept Tactical Systems Repairer
	33V	Electronic Warfare/Intercept Aerial Sensor
	33Y	Strategic Systems Repairer (EW Tactical Systems Repairer)
	37F	Psychological Operations Specialist
	38A	Civil Affairs Specialist
	46Z	**Combined**
	46Q	Journalist
	46R	Broadcast Journalist
	51T	Technical Engineering Specialist
	55R	Ammunitions Stock Control and Account Specialist
	71C	Executive Administrative Specialist
	74B	Information Systems Operator
	81C	Cartographer
	81L	Printing and Bindery Specialist
	81Q	Terrain Analyst
	81T	Topographic Analyst
	82D	Topographic Surveyor
	93B	Aeroscout Observer
	96B	Intelligence Analyst
	96D	Imagery Analyst
	96F	Psychological Operations Specialist
	96U	Unmanned Aerial Vehicle Operator

The 17 Family Second-Tier System

Cluster	MOS	Job Title
	97B	Counterintelligence Agent
	97E	Interrogator
	97G	Multi-Discipline Counter Intelligence
	97L	Translator/Interpreter (RC)
	97X	Linguist
	98C	Signals Intelligence Analyst
	98G	EW Signal Intelligence Voice Interrogator
	98H	Morse Interceptor
	98Z	**Combined**
	98D	Emitter Locator/Identifier
	98J	Noncommunications Interceptor/Analyst
	98K	Non-Morse Interceptor/Analyst
	98X	EW/SIGNIT Specialist

Skilled Technical (ST) 3

Cluster	MOS	Job Title
17	54B	Chemical Operations Specialist
	82C	Field Artillery Surveyor
	93C	Air Traffic Control (ATC) Operator
	93P	Flight Operations Coordinator
	95B	Military Police
	95C	Corrections Specialist

Note: The constrained H_d clustering algorithm was used, with the resulting families <u>moderately</u> modified by judgment.

Evaluation of Sets of Families

Table 3 shows MPPs for five sets of families separately for: (1) families clustered by constrained H_d modified by judgment; and (2) families clustered by the unconstrained H_d empirical algorithm. The 10 operational aptitude area (AA) families, as noted, were not subjected to the empirical clustering technique or to modification by judgment.

Table 3

MPP for 12 Conditions

| Family Size | Constrained H_d Adjusted by Judgment | | | | Unconstrained H_d | | | |
| | Biased | | Unbiased | | Biased | | Unbiased | |
	MPP	SD	MPP	SD	MPP	SD	MPP	SD
150	.334	.012	.195	.013	.334	.012	.195	.013
17[1]	.186	.012	.145	.012	.216	.012	.193	.012
17[2]	.201	.012	.165	.013	.216	.012	.193	.012
13	.168	.023	.138	.021	.187	.048	.172	.043
10	.149	.015	.123	.012	.183	.026	.170	.023
AA[3]	.023	.012	.023	.012	--	--	--	--

[1] Clustered by constrained H_d first and moderately adjusted by judgment to better conform with operationally-based job families.
[2] Clustered by constrained H_d first and only slightly adjusted by judgment.
[3] Aptitude Areas used with operational families.

The optimal allocation of individuals to jobs or families was constrained in all simulations to conform proportionately to the actual operational distribution of enlistees to jobs in 1989, the most relevant year. For example, the infantryman MOS, 11B, would receive 12.5 percent, or 2,500, of the 20,000 enlistees assigned to the 150 jobs. Appendix C1 shows the percentages used in making weighted operational assignments to each MOS. Appendix C2 shows the percentages for each job family.

Examining first the unbiased, unconstrained H_d results for families in which H_d has been maximized, we see a very small gain in MPP between the 10 and 13 job families and a greater increase between 13 and 17 and between 17 and 150 families. Note the relatively large SD for the 13 family set. Using the operational

AAs with an MPP of .023 as the base line, the unconstrained H_d algorithm yields unbiased MPPs that are more than eight times greater for all family sizes.

When we examine the MPPs for the constrained H_d solution modified by judgment, we find, for the unbiased condition, MPPs steadily increasing from .123 for the 10 family LSE set to .195 for the 150 family set. The families created using constrained H_d modified by judgment is of most interest in these comparisons because they reflect, in general, clustering consistent with judgment that is more likely to be acceptable for operational use. At the same time they retain much of the CE obtained by using unconstrained H_d. Note that in examining the two sets of 17 families, the set where H_d was only slightly adjusted by judgment provides an unbiased MPP gain of .02 over the 17 family set that was moderately adjusted to conform to existing CMF and AA families. This difference is further evaluated in Table 6 below.

Note that the 150 family set is identical for the unconstrained H_d and the constrained H_d conditions because no changes in the structure of the set was made. In both conditions, the 150 families consist of 146 single MOS families and 4 multi-job families. The results of Table 3 begin to shape our choices for the family structures to employ in the two-tiered system.

To better comprehend the results found for the operational AAs, we note that each trial set of families employ unbiased least squares estimates (LSEs) of predicted performance using all nine ASVAB tests whereas the AAs use unit-weighted, four-test composites. It is clear in comparing AAs with any one of the trial sets that the trial sets are always superior. The use of weighted composites of all ASVAB tests in conjunction with new job families result in very large improvements in MPP over the existing AA system. This finding has been corroborated by the authors and their colleagues in all previous research using different data sets.

In Table 4 we examine the relationship between the existing 10 operational AAs and a trial set of 10 families employing LSEs based on the full set of ASVAB tests. We see the clear superiority of the unconstrained H_d solution in forming a new set of 10 families and at the same time "best" weighting the 9 ASVAB tests (an unbiased MPP of .170 for the LSEs vs an MPP of .023 for the operational families).

Table 4

Comparison of Mean Predicted Performance for 2 Unconstrained H_d
Assignment Conditions (10 H_d Job Families) with the AA job families:
Classification Effects Only

Number of Families	MPP	SD
10 LSEs (biased)[1]	.183	.026
10 LSEs (unbiased)[1]	.170	.043
10 AAs operational (unbiased)	.023	.012

[1] Job families using H_d LSE composites (unconstrained).

If we next compare the results for the operational set of 10 families using LSEs as AVs with the base-line condition shown in Table 5, we again find clear superiority of the set of LSEs (an unbiased MPP of .123 vs .023). While the set of 10 LSEs using the operational families clearly provide a major improvement over the operational AAs, its use for the second tier system appears less promising on the basis of MPPs than LSEs for either the sets of 13 or 17 families. Since the set of 17 families produce high MPPs, we focus further on them.

Table 5

Comparison of Mean Predicted Performance for 2 Constrained H_d
Clustering Conditions and the Operational Condition (10 Operational
Job Families): Classification Effects Only

Number of Families	MPP	SD
10 operational LSEs (biased)[1]	.149	.015
10 operational LSEs (unbiased)[1]	.123	.012
10 AAs operational (unbiased)	.023	.012

[1] Operational job families using LSE composites

In Table 6 the two sets of 17 families are compared using LSE weights that include both positive and negative weights with LSE weights that are all positive because the tests that do not provide positive weights have been removed before the computation of regression weights. Using both types of weights, the 17 family set that is only slightly modified from a constrained H_d solution results in higher MPPs than the more adjusted set of 17 families (.165 vs .145), as was shown earlier in Table 3. The slightly modified 17 family

set is not shown. When only positive weights are employed there is only a very slight reduction in the MPPs for both sets (.164 vs .142). Since the second tier would be employed largely for counseling purposes and for establishing minimum cut scores, rather than for actual job matching, only positive weights are suggested for use in the second-tier composites. Also, since the second tier composite scores are visible to both counselors and enlistees, it seems more desirable to use clusters that are more homogeneous and rational. Given the intended use of the second tier and the importance of user acceptance of the visible system, the more highly modified set of 17 families that is consistent with the operational CMF system and AA families is provisionally considered more desirable. This is the 17-family set shown in Table 2. Thus at this point in our analyses, a transparent or invisible first-tier system of 150 families along with a visible second-tiered system of 17 constrained families (with positively weighted LSEs) are selected for more detailed consideration. As shown in Table 6, these sets have MPPs of .195 and .142 (for positive weights) respectively. Table 7 summarizes the key comparisons among the family sets.

Table 6

Unbiased Mean Predicted Performance for 2 Assignment Conditions by Number of Job Families: Classification Effects Only

	All Weights		Positive Weights	
Number of Families	MPP	SD	MPP	SD
17^1	.145	.012	.142	.013
17^2	.165	.013	.164	.012

[1] Clustered by constrained H_d first and then adjusted on the basis of judgment
[2] Clustered by constrained H_d first and then adjusted slightly on the basis of judgment (relying heavily on H_d)

Table 7

Unbiased Mean Predicted Performance for 4 Assignment Conditions by Number of Job Families: Classification Effects Only

Number of Families	MPP	SD
150	.195	.013
17^1	.145	.012
13^1	.138	.021
10^2	.123	.015

[1] Clustered by constrained H_d first and then moderately adjusted on the basis of judgment.
[2] Operational LSE job families.

Imposing Quality Constraints

In considering the new set of 17 families for the second tier, it may be desirable to not only use positively weighted LSEs but to constrain the optimal distribution of enlistees to jobs so that the MPP for each family is positive. In other words, after allocation of individuals to jobs, all MPPs would have values above the statistical standard score of zero. This requires imposing a constraint on the optimal allocation algorithm resulting in a lowering of the overall MPP in order to remove or largely remove negative family MPPs. Thus, school proponents of some families would not feel that less able enlistees were being assigned to families of interest to them when, in fact, predicted performance in all families is being raised over that provided by selection alone.

In Table 8 we show the results of two levels of constraints on the H_d 17 family and operational family sets: (1) using only positive LSE weights and (2) constraining the linear program (LP) assignment program to limit the size of negative MPPs for each family. Note for the 17 family set that except for the Clerical (CL) 1 family that still has a very small negative MPP of -.016 remaining, all other MPPs are positive. Note, too, that there is a sizable reduction in the overall MPP from .142 to .106 to achieve the desired condition of eliminating negative MPPs for families. This result demonstrates that an LP program can be used to improve MPP through allocation of personnel even when constraints are imposed. It should be kept in mind that a statistical standard score value of zero in this investigation is equal to a mean AA score of about 105. This example further demonstrates that it is practical to impose operational constraints on the optimal allocation algorithm to achieve various policy objectives. For example, it may by possible to constrain the LP algorithm so that it would not permit the occurrence of a negative MPP for one particular family and this in turn may result in only a minimal reduction in overall MPP.

Table 8

***Comparison of MPP for the Operational AAs and 17 families Based on
Constrained H_d and Moderately Modified by Judgment Using Optimal
Assignment and Further Constrained to Avoid Negative MPPs***

Operational Families			Constrained 17 Families		
Family	MPP	SD	Family	MPP	SD
Overall	0.014	0.025	Overall	0.106	0.014
CL	0.224	0.071	CL1	**-0.016**	0.022
CO	0.268	0.031	CL2	0.099	0.003
EL	**-0.932**	0.052	CO1	0.075	0.003
FA	0.106	0.084	CO2	0.059	0.005
GM	**-0.408**	0.094	EL1	0.029	0.009
GT	**-1.020**	0.477	EL2	0.016	0.009
MM	0.155	0.063	EL3	0.043	0.008
OF	**-0.107**	0.061	FA	0.002	0.007
SC	0.386	0.090	GM1	0.870	0.219
ST	0.019	0.044	GM2	0.065	0.008
			MM1	0.390	0.050
			MM2	0.152	0.175
			OF	0.060	0.004
			SC	0.057	0.006
			ST1	0.077	0.006
			ST2	0.051	0.057
			ST3	0.016	0.007

Note: Negative family MPPs are in bold print.
[1]   Using all positive LSE weights

49

In sharp contrast to the results obtained with the set of 17 families are the results shown in Table 8 for the operational AAs. Four of the ten families have large negative MPPs ranging from -1.02 to -.107. We could not find a feasible solution that would remove all of the negative MPPs. The existing system, then, produces only a very moderate gain in overall MPP (.014), and at the cost of sending relatively poorly matched soldiers into a number of job families.

Relationship between Number of Families and MPP

In Table 9 we examine more closely the increase in MPP as the number of job families are increased. In the previous analysis, we examined only four family sizes (10,13,17, and 150). Here we examine 22 sets of family sizes, ranging from 4 to 150 for the unconstrained H_d algorithm. By not considering in this analysis constrained H_d clusters that are modified by judgment, we can better evaluate the effect of number of job family sets on MPP.

Biased MPPs continue to increase from 4 through 150 (from .1248 to .3337). See Figure 3 for a graphic representation of these values. The increase in biased MPPs does not start to asymptote even at 150 jobs. In contrast, the unbiased MPP continues to rise only to about 66 families and then asymptotes (or slightly declines). See Figure 4 for a graphic representation of these values. The shrinkage between biased and unbiased estimates of MPP continues to increase as the number of families increase. See Figure 5 for a graphic representation of these difference values. For example, the MPP difference between biased and unbiased shrinks about .14 for the 150 families and .08 for the 66 families. It is apparent that the opportunity to fit error in the biased sample increases as the number of families increases.

For the 150 job-family size, the unbiased MPP is .1950 compared to .2120 for the highest MPP value found for the 66 families, a difference of .017. Also the three families between the 150 and 66 (85, 104, 127) have slightly higher MPPs than does the 150 families, with MPP differences ranging from .015 to .008. We initially believed that one way to reduce the shrinkage in MPPs was to eliminate from the analysis smaller MOS sizes. We first reduced the number of MOS from 150 to 127 and then we reduced the 127 set to 104 by eliminating the smaller sample-size MOS until the family sizes were equal to the 127 and 104 sets used in the original analysis. In effect, for the set of 104 families, we were comparing families formed entirely by the H_d algorithm with families partially selected by sample size. For the set of 127 families, we were comparing families selected by H_d with families entirely selected by sample size.

50

Table 9

MPPs for Unconstrained H_d Families Ranging in Size from 4 to 150

Number of Families	Biased		Unbiased	
	MPP	SD	MPP	SD
4	.1248	.0118	.1207	.0119
5	.1326	.0284	.1283	.0277
6	.1445	.0120	.1387	.0121
7	.1476	.0221	.1406	.0208
8	.1709	.0246	.1597	.0228
9	.1781	.0122	.1662	.0124
10	.1828	.0256	.1697	.0233
11	.1896	.0122	.1758	.0123
12	.1976	.0123	.1809	.0125
13	.1874	.0475	.1720	.0427
14	.1927	.0490	.1756	.0438
15	.2102	.0122	.1882	.0122
16	.2111	.0122	.1886	.0122
17	.2155	.0122	.1931	.0121
21	.2243	.0121	.1979	.0122
25	.2349	.0121	.2032	.0121
40	.2675	.0120	.2118	.0127
66	.2950	.0119	.2120	.0123
85	.3067	.0118	.2104	.0124
104	.3142	.0119	.2068	.0129
127	.3256	.0118	.2025	.0129
150[1]	.3337	.0118	.1950	.0134

[1] Contains 146 families that are single MOS families and 4 families that have several MOS each.
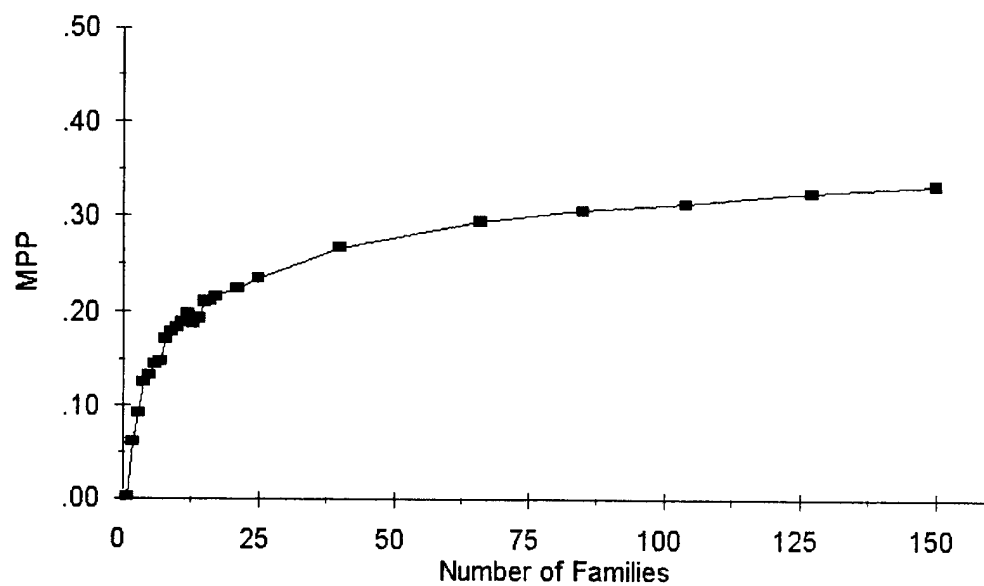
Figure 3 Biased MPPs by Family Size
(Families based on H_d)
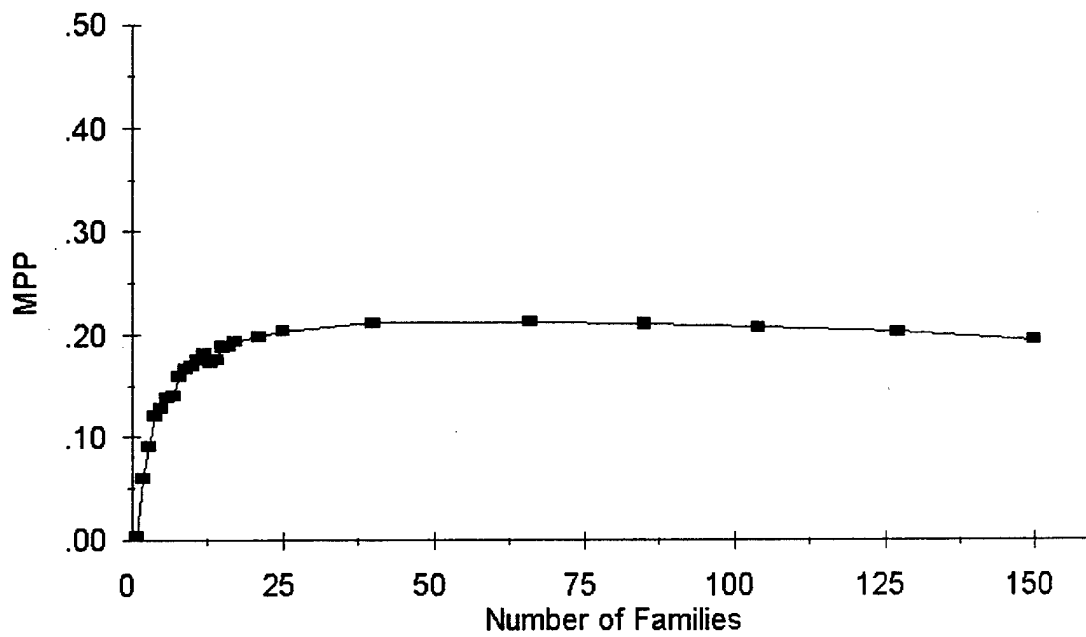
Figure 4 Unbiased MPPs by Family Size
(Families based on H_d)

Figure 5 Biased/Unbiased MPPs by Family Size
(Families based on H_d)

Table 10 shows the superiority of the unconstrained H_d method. Obviously in forming families based on eliminating smaller sized MOS, we were also changing the overall structure of the families. Some of the smaller MOS that were eliminated may ha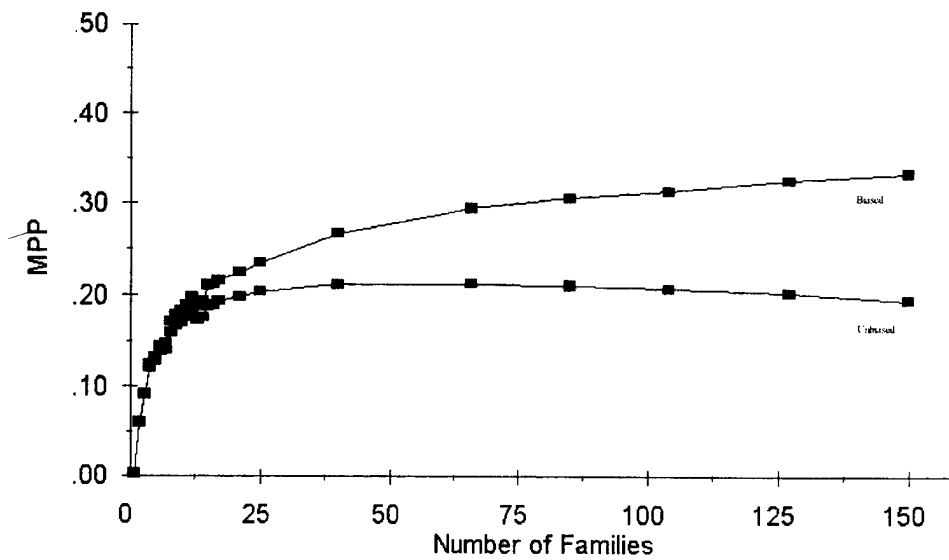ve also been more differentiated from other MOS than larger sample MOS remaining in the analysis. A more refined experiment would be required that controlled for both family structure and size to obtain a better understanding of the effect of reducing the number of families using a criterion of analysis sample size. The analysis presented in Table 10 confirms the value of the H_d algorithm in forming families when the objective is to maximize CE.

Table 10

Comparison of Unbiased MPPs for Two Family Sizes Formed by Two Methods

	H_d			Families Limited by Size[1]	
Family Size	MPP	SD		MPP	SD
104	.207	.013		.192	.013
127	.203	.013		.196	.014

[1] Note. Four families were comprised of more than one MOS in each family.

Consistency of Results

In previous research (Johnson & Zeidner, in preparation) 66 MOS were analyzed to demonstrate the advantage of a two-tiered system over the operational system. In Table 11 we compare four family sizes obtained in this earlier research with the closest family size equivalents employed in the present research. It should be noted that most of the 66 MOS were also embedded in the larger 150 MOS investigation. Such a comparison, then, should provide a general index of consistency of findings across overlapping investigations. All family sizes other than the AA operational families employed LSE weights in the assignment composite.

Table 11 shows that the 150 families produced an overall MPP that is comparable to the MPP for the 66 families of the earlier investigation (.195 vs .190). The comparison of 17 families (adjusted slightly by judgment from the constrained H_d algorithm) also produced very similar results (.165. vs .161). The 10 families and AA families comparisons produced, as expected, larger differences of MPP around .02 between the two investigations. The larger number of MOS in the "150 family investigation" result in relatively heterogeneous families when the number of families permitted is restricted to 9 or 10 families. In contrast,

the smaller number of MOS in the "66 family investigation" produce more homogeneous families when the number of families is restricted to 9 or 10.

Table 11

Comparison of Investigations Using 150 and 66 Job Families

"150 Investigation"		"66 Investigation"	
Size	MPP	Size	MPP
150	.195	66	.190
17^1	.145		
17^2	.165	16	.161
13	.138		
10	.123	9	.149
AA	.023	AA	.047

[1] Clustered by constrained H_d first and moderately adjusted by judgment.
[2] Clustered by constrained H_d first and only slightly adjusted by judgment.

Selection of the Two-Tiered System

The First-Tiered System

Although the 150 family set results in a slightly lower MPP (.195) than do the job family sets between 66 and 127 (with MPPs ranging from .212 to .203), we have selected the 150 job family set for the first tier for the reasons indicated below:

1.   For the 150 family set it is not necessary to defend the logic of placing a given set of MOS in the same family since the system is comprised of 146 single job families. There would be no confusion over cluster homogeneity in the first-tier system.

2.   There is greater flexibility in making changes in assignment variables (AV) for single MOS families than for multiple job families in future investigations intended to evaluate new MOS or modify changes in existing MOS.

3.   There should be greater management credibility for the use of single-family clusters than for 40 or 66 family clusters based on H_d. While the formation of clusters by H_d maximizes MPP, the resulting clusters or families often lack rationality.

56

4. The 150 first-tier system is completely compatible with the visible second-tier 17-family system since all MOS remain in consistent groupings.

5. Differential assignment theory supports the use of a larger number of families with stable weights producing higher MPPs. More stable families produce larger mean validity and lower intercorrelations, plus greater optimization efficiency.

6. The biased MPP values may be closer approximations to population values as MOS sample sizes increase beyond the sample sizes used in the present investigation.

The Second-Tiered System

The 17-family set first formed by the constrained H_d clustering algorithm and then moderately modified by judgment is selected for the second-tiered system. This set produces an MPP of .142 using only positively weighted LSEs. Although lower than the MPP of .195 of the first-tiered set, the 17-family set retains much of its validity for its intended purpose of counseling recruits in the job assignment process, for utilizing cut scores and for keeping a record helpful to future personnel decisions. Further, the second-tier system is largely consistent with the CMFs currently used by the Army in managing the entry-level job structure and is also consistent with the current AA system. No pair of MOS that are together in the 17 family system fails to be together in the operational AA system (except for GT). Four of the nine AAs in the 17-family set are divided into two relatively homogeneous sets of clusters, two AAs are divided into three sets of clusters, and three remain as single clusters. See Table 2 for a description of the 17 family set and the MOS that compose each family.

Additionally, the 17-family system should not be overly burdensome to use for record-keeping purposes in place of the current AAs. Thus, the visible portion of the two-tiered system will correspond much more closely to the MPP values of the first tier; have much more accuracy than the current AAs; and not change the current AA designators for any given MOS. In summary, the second-tiered system should appear familiar to the operational community and also be user-friendly, especially after the transformation to the Army conventional standard score scale (ACSS).

MPP Gains Attributable to the Two-Tiered System

MPP Gains for the First Tier

Table 12 shows the gains in making optimal assignments for various sized families. Using the 150-family set as yielding the maximal obtainable MPP of 100%, we see that the families of other sizes yield gains of 74% down to about 12% for the operational aptitude areas.

Table 12

Comparison of 150 Job Family Composite Yielding Maximally Obtainable MPP with Other Composites: Classification Effects Only

Strategy	MPP[1]	% Retained
150 job families (LSEs)	.195	100
17 job families(LSEs)[2]	.145	74.4
13 job families (LSEs)[2]	.138	70.8
10 operational job families (LSEs)	.123	63.1
10 operational job families (AA)	.023	11.8

[1] Unbiased estimates
[2] Clustered by constrained H_d first and then moderately adjusted on the basis of judgment.

Comparing the Independent Contributions of Selection with the New Two-Tiered Classification System

In examining Table 13 we see that the independent contribution of classification to total MPP is greater than the independent contribution of selection (53.9% vs 46.1%). The selection computation is based on the weighted (for accessions) average validity of .607 for AFQT against SQT for 170 MOS. An acceptance rate of 85 percent on AFQT is used for the selection ratio. This type of comparison is not well known because simulations of the classification process based on empirical data are rarely undertaken. However, Table 13 clearly points to the benefits that could be achieved by employing a two-tiered system in place of the current AA system.

Table 13

Independent Contributions of Selection and
Classification Effects to MPP

Condition	MPP	Percent of Total MPP
Selection	.167	46.1
Classification	.195	53.9

Transforming ASVAB Test Scores for the New Two-Tiered System

Transforming ASVAB Test Scores into Modified Statistical Standard Scores (Tier 1)

The scale proposed for use in the operational context would have a mean of zero and varying SDs for each of the 150 assignment variables and would provide the MPP results obtained in our simulations of the first tier. In accomplishing the first-tier simulations, we converted the operational ASVAB scores to modified statistical standard scores. This scale has a mean of zero across all AVs, but variations in the SDs proportional to test validities for individual AVs. In order to avoid transforming the operational test scores to standard scores, the beta weights were transformed to "u" weights for each of the 150 assignment composites and then a constant "k" was subtracted from each composite. The resulting transformed scale provides the same estimate of benefit, .195, as in the simulation experiment and at the same time permits the convenient use of operational ASVAB test scores. Since the computational process and intermediate results are transparent to the users, the use of best weights (positive and negative LSEs), rather than the use of only positive weights (and a reduced number of tests), is employed. The procedure for transforming scores is outlined in Appendix D1. Appendix D2 provides the beta weights used in simulation for the composites for the first tier and Appendix D3 provides the transformation weights "u" and constants "k".

Transforming ASVAB Test Scores into an Army Conventional Standard Score (ACSS) Scale (Tier 2)

The ACSS proposed for use in the operational context, with an expected mean of 100 and SD of 20, is uniformly provided for all AVs. The use of ACSS results in a minimum change in the current operational system as a consequence of replacing unit-weighted composites with LSEs. Thus, a given score has the same meaning and use as provided by the conventional score scale of the current operational system. ARI researchers thought it was advisable to retain the use of the traditional aptitude area scoring scale that is tied back to the youth population parameters. It was recognized that in adopting the ACSS all means and

59

standard deviations would be equivalent and this, in turn, would lower MPPs since hierarchical effects (the effects of using variations in validity of AV composites) would be removed. In a simulation of the ACSS, we found an MPP of .097 (with positively weighted LSEs) as compared to the statistical standard score scale MPP of .142. The procedure for transforming the operational ASVAB test scores into ACSS scores for use in the second tier is given in Appendix E1. Appendix E2 provides the beta weights for the set of tests yielding positive weights. This is highly desirable since these weights are visible. Appendix E3 provides the transformation weights "u" and the constants "k."

Small Entry-Level MOS

There are about 80 small input entry-level MOS for which there are no data. Consequently, these MOS were attached to MOS kernels used in the first tier. Attachments were made on the basis of previous research findings and judgment. The small entry-level MOS would employ the same AV as for the kernel to which it was attached. The AVs for MOS kernels are based on empirical data. Appendix F lists the small entry-level MOS and the larger MOS (kernels) to which they are attached.

Findings of the Supplementary Experiment: The Two Types of Test Takers

Problem

In preparing the data set used in this second investigation, we initially searched the files for subjects that might have taken the same SQT exam twice in the same year. In such instances, we dropped the subjects from analysis since our data did not permit distinguishing between the first SQT and the second SQT. Later in the analysis we realized that subjects included in our data set could have taken SQT exams across different years, although they would be taking different versions of the exam each year. Army policy during the period that the data for our analysis was collected directed that enlistees take the SQT exam for their MOS regardless of how well they previously performed on the exam. Thus our data set could have included both individuals who had taken the SQT exam for the first time, after about a year of service in the Army, and at the same time the data set also could have included individuals who had taken different versions of SQT exams as many as three times. If this were found to be the case, we needed to be concerned about the effect of experience during the first tour of duty on our results.

We were well into evaluating the central investigation, the two-tiered investigation, when we received analysis results that we requested from ARI researchers (since they had access to social security numbers for identification purposes) showing that about half of the subjects in our data set had taken SQT exams two or three times during their first tour of duty. Although initial analyses showed that there were no quality differences (e.g., first-time test takers had equivalent ASVAB scores to multiple-time test takers) the data also showed that additional experience on the job raised SQT scores. This, in turn, could affect the relationships between predictors and criteria so that regression weights could be different when computed for the total group and for the subsample of first-time test takers.

To investigate the question of experience on the results of our central experiment, a new experiment was undertaken to compare effects for a variety of conditions. Ideally, it would have been desirable to conduct the central experiment with first-time test takers only. This would be a more homogeneous group in terms of experience. However, it would have reduced the A, B, and C samples in half, precluded the inclusion of many MOS with smaller sample sizes and created less stable assignment variable weights. The present research was undertaken to answer the following specific questions: (1) Do the LSEs computed in the central investigation for use as AVs result in higher, the same or lower MPP than the AVs we would obtain by reducing our analysis sample in half by discarding multiple-times test takers from our analysis samples? (2) What is the overall effect of the inclusion of experienced cases in the analysis sample?

Method

We identified 11 MOS in the present investigation with each MOS representing one of nine operational families. However, two Combat Arms (CO) MOS were included because of the large proportion of enlistees assigned to CO MOS and two Operators and Food (OF) MOS were also included because of the diversity of the two groupings of jobs in this family.

Table 14 shows the 11 MOS selected, all from FY 1988, and the subsample sizes for each MOS. The size for the total sample is 21, 565; first-time test takers comprise 54% (11,719) of the total sample and multiple-times test takers comprise 46% (9,846). The first-time sample is composed of subjects who have taken the SQT entrance exam once. The multiple-times sample is composed of subjects who have taken the SQT exam two or three times, using fiscal 1986 as the earliest starting year for our samples. As noted earlier, a different version of the SQT exam is given each year for each MOS and each enlistee is required to take the SQT exam each year during her/his first tour of duty.

The overview of the experimental design is shown in Table 15, along with our hypotheses, which were made before data analysis, for 4 experimental conditions. The analysis sample equals one-half of the first-time cases and all of the multiple-times cases. The weights for 11 AVs are computed on four separate samples: first-time, multiple-times, total cases and modified total cases. The evaluation sample equals a random half of the first-time cases for the 11 selected MOS. This sample is not further subdivided. The cross (simulation) sample equals 30,000 first-time cases drawn from MOS other than the 11 selected MOS. This sample was divided into 30 (instead of the usual 20) replication subsamples to permit the obtaining of more credible critical ratios across overall means for the 4 experimental conditions.

The typical ABC (analysis, evaluation cross samples) design, omitting the double cross validation elaboration, is utilized. AV weights are computed in sample A for each condition, evaluation weights are computed in sample B, and both the AV and the EV weights are applied to ASVAB scores for cases (entities) in the cross sample (C), followed by optimal assignment of the cases in sample C to one of the 11 selected MOS. Equal quotas for the 11 MOS are used.

The focus of our analysis is on the difference in MPP between the smaller number of first-time cases and the cases in the total sample. We have little interest in the difference in means between experienced cases and total cases and a secondary interest in the difference between the means of first-time and experienced cases.

Table 14

Sample Sizes for Comparing First and Multiple Time SQT Test Takers

MOS	Name	Total Sample	First-Time SQT Sample[1]	Multiple-Times SQT Sample[2]	Year	Aptitude Area
11H	Heavy Anti-Armor Weapons Infantryman	2,470	1,251	1,219	88	CO
13F	Fire Support Specialist	1,848	1,037	811	88	FA
19K	M1 Abrams Armor Crewman	2,810	1,143	1,667	88	CO
31C	Single Channel Radio Operator	2,926	1,020	1,906	88	SC
31V	Unit Level Communications Maintainer	1,798	1,074	724	88	EL
52D	Power Generator Equipment Repairer	2,537	1,177	1,360	88	GM
67V	Observational/Scout Helicopter Repairer	793	375	418	88	MM
75B	Personnel Administration Specialist	1,647	873	774	88	CL
91K	Medical Laboratory Specialist	590	331	259	88	ST
92G	Food Service Specialist	3,270	2,983	287	88	OF
16S 1	Man Portable ADS Crewmember	876	455	421	88	OF

[1] First-time sample is comprised of subjects who have taken the SQT exam once. These subjects are required to take the exam each year of their first tour of duty.

[2] Multiple-times sample is comprised of subjects who have taken the SQT exam two or three times, a requirement for all enlistees during the first tour of duty.

The objection to the use of multiple-times cases (experienced cases) hinges on the supposition that "truth" is best represented by the use of only first-time cases, a more homogeneous group. This design temporarily accepts this assumption of truth and provides a test bed of first-time cases in which it can be determined whether multiple-times cases detract from classification efficiency (CE) present in a set of AVs. The possibility that the operational AVs being proposed have less CE because of the inclusion of multiple-times cases in the ABC sample is a serious issue that is addressed.

Table 15

Overview of the Experimental Design

Condition	Sample[1] AV	EV	Cross	Hypothesis[2]
1	All multiple-time cases from MOS 1 - 11 (Multiple-time takers)	1/2 first-time cases from MOS 1 - 11	All first-time cases from previous cross analyses, except cases from MOS 1 - 11	This condition should produce higher MPPs than condition (2).
2	1/2 first-time cases not used in EV from MOS 1 - 11 (First-time takers)	same first-time cases as in (1) above	Same as cross in (1) above	This condition should produce lower MPPs than all other conditions.
3	All multiple-time cases from MOS 1 - 11 plus 1/2 first-time cases not used in EV as in condition (2) above	Same first-time cases as in (1) above	Same as cross in (1) above	This condition should produce MPPs higher than (1) and (2) above.
4	Same as in (3) above, but SQT scores of first-time test-takers corrected for experience			This condition should produce the highest MPPs.

[1] First-time sample is comprised of subjects who have taken the SQT exam once. These subjects are required to take the exam each year of their first tour of duty. Multi-times sample is comprised of subjects who have taken the SQT exam two or three times, a requirement for all enlistees during the first tour of duty.

[2] If Condition 3 and 4 produce MPPs equal to or higher than Condition 2, the use of multiple-times cases has no negative impact on result. In fact, the use of a larger sample size may be beneficial in providing more stable weights.

Results and Conclusions

In Table 16 is presented the MPPs for the 4 conditions in the cross sample. Under the assumption that Condition 1 provides accurate measures of the criterion and that the relationships among AVs provide the most differential validity, because of the lower intercorrelations among the AVs (r), the overall MPP for this condition should be higher than for Condition 2. Condition 1 employs only multiple-times test takers as AVs; the estimate of job proficiency for these experienced cases should be relatively accurate. At the same

time, experienced individuals should be employing a range of specialized skills appropriate for their jobs. The MPP is .210 for Condition 1.

Table 16

MPPs for Four Experimental Conditions in Cross-Sample (N = 30,000)

	Cross Sample	
Condition	MPP	SD
Multiple-Times	.210	.014
First-Time	.183	.016
Total Sample	.223	.016
Corrected Total Sample	.214	.016

Condition 2 uses only first-time test takers for the three samples (AV, EV and Cross). Because of the relative homogeneity of the sample, we provisionally consider Condition 2 to be the truth condition, as mentioned earlier. However, this condition provides a more g-loaded prediction of the criterion than Condition 1, resulting in a higher r. The MPP for this condition is .183 as compared to .210 for Condition 1, as expected. We find that in this research, consistent with all previous investigations, that r makes a greater contribution to MPP than the average validities of the AVs (R).

To the extent that Condition 1 has a higher MPP than Condition 2, the total sample condition, Condition 3, should have a higher MPP than Condition 2. Assuming that the first-time sample does not provide poor information, the addition of a relatively large number of first-time cases to the sample of multiple-times cases should compensate for the increase in g-loaded AVs of first-time cases and thus provide an equal or improved prediction of the criterion. Thus, the MPP for Condition 3 should equal or exceed Condition 1.

However, if the first- and multiple-cases do not equally predict the criterion using different weights for the ASVAB tests reflecting poorer information provided by the first-time cases, the addition of first-time test takers would decrease MPP. Thus the MPP for Condition 3, while exceeding that of Condition 2, would be less than for Condition 1. The MPP found for Condition 3 is .222, exceeding both of the other two conditions, as expected.

Because Condition 3 is higher than Condition 2, we are completely justified in the inclusion of multiple-times test taker cases in the design of the central investigation. The critical ratio (CR) of the MPP difference of .039 between the two conditions is statistically significant. This CR considers the correlation among the 30 MPPs obtained for a pair of conditions that exceeded .90 for all pairs of conditions.

In Condition 4 we correct the SQT scores of first-time test takers for the mean SQT differences between them and multiple-times test takers for each MOS. The correction in the criterion scores is intended to equate the two groups. We would expect MPPs for Condition 4 to be slightly higher than the MPP for Condition 3. However, the MPP for Condition 4 is .214, or .008 less than for Condition 3. While the critical ratio is statistically significant, it is of no practical significance. This justifies not correcting for experience differences in the central investigation.

In Table 17 we explore the internal relationships among MPPs, R, and r for the four conditions to gain further insight of the results obtained from the second investigation. We see from the table values of R and r both in the cross and back samples that are generally close. We note that the condition that has lowest r has the highest MPP in both the cross and back samples, as expected. In Condition 2 for first-time test takers in the cross sample, we see the highest intercorrelations among the AVs (.887) and also the lowest MPP (.183), as expected. Conditions 3 and 4 produce R, r, and MPPs that are very similar. The results are generally consistent with Brogden's (1959) formulation of MPP $= f_{(m)} R \sqrt{1 - r}$ computed in the AV (back) sample, where m is the number of jobs, R is the mean validity, and r is the mean of the intercorrelations among AVs (predicted performance).

Table 17

**Relationship among R, r, and MPPs for Four Conditions**

Condition	Cross Sample			Back Sample (AVs)[1]		
	R	r	MPP	R	r	MPP
1	.520	.879	.210	--	--	--
2	.553	.887	.183	--	--	--
3	.530	.875	.223	.537	.878	.297
4	.532	.800	.214	.541	.884	.293

[1]   Values were not computed for Conditions 1 and 2 in the back sample because different subsamples were used. MPPs in the back sample were computed by Brogden's formulation, $f_{(m)} R \sqrt{1 - r}$.

In summary, then, the results comparing experienced and inexperienced enlistees during the first tour of duty clearly indicates that the weights obtained for the two-tiered system in the central investigation can form the basis of our conclusion and recommendations.

Discussion of Findings

Job Families and Classification Efficiency

Results of the present research continue to support an early differential assignment theory principle that maximum MPP is obtained by utilizing the LSEs of the criterion as AVs for all jobs having adequate or stable validity data. The H_d clustering algorithm, starting with 170 MOS, created a set of 150 job families for the first tier system composed of 146 single MOS and 4 families consisting of several MOS each.

The 150 families were considered distinct and stable for initial consideration as the first tier system. The 150 MOS family structure is nearly two and one half times the number of MOS (66) used in the earlier demonstration investigation of the two-tiered system (Johnson, et al., in preparation). The present research incorporates all but about 80 smaller sized entry-level MOS in the data set used to compute AVs. In all, about 90% of enlistees entering the Army are assigned to one of the 150 job families included in the first tier structure prior to making attachments by judgment. The sampling of 150 job families having complete ASVAB and SQT data for 260,000 enlistees constitutes a comprehensive representation of the Army's selection and classification system--a data set that in all likelihood would be very difficult to once more assemble in the foreseeable future.

The unbiased overall MPP for the 150 job family structure is .195 compared to the MPP for the existing AA system of .023--a gain of more than eight times in classification efficiency. A smaller gain of MPP associated with an increase in m (number of families) is shown in the present investigation as compared to gains shown in investigations using synthetic scores. The smaller gains are attributable to research conditions including: (1) a very stringent unbiased research design; (2) the use of empirical rather than synthetic scores; (3) longitudinal rather than concurrent validity; and (4) data corrected to the Army input population rather than to the youth population.

The MPP for the previous demonstration investigation using a 66 family structure is about the same as for the present investigation of 150 families (.190 vs .195). Initially, this was a surprising result to us. However, a closer analysis of the family structures in the two investigations showed an inclusion of a larger proportion of technical MOS (with higher validities) and clusters that were more distinct from one another in the "66" investigation compared to the "150" investigation. In the present research, using 150 MOS, the overall job structure was not as differentiated, with many jobs located at the borders between clusters. Note, too, that in the experiment in which we evaluated the effect of experience on weights used to compute LSEs, we included 11 MOS representing 9 different AAs and obtained an MPP of .214 as compared to the MPP of

.195 in the 150 MOS investigation. Again, this demonstrates the effect of distinctiveness and overall family structure on MPP.

In our analysis we were also interested in finding differences attributable to the use of the H_d algorithm using all MOS with moderately sized validity samples to define job families compared with an alternative H_d method that utilized only large sample sizes in conjunction with H_d to form job families. The H_d model, using the maximum number of job families at the expense of the inclusion of relatively small analysis samples for the computation of some AVs, resulted in higher MPPs, again demonstrating the value of Horst's formulation for maximizing the CE differential using LSEs of the criterion as AVs in an optimal assignment system, even when some validity samples were smaller than is generally considered to be necessary for the provision of stable weights for LSEs.

Based upon the previous demonstration investigation of 66 families, the authors believe that several alternative sets of job families would be effective for use as AVs in the second tier system. Of these, the least satisfactory alternative would be the use of the existing 10 operational families (including the GT AA). The shredding of operational families into 12 to 19 families, and further modified considering currently employed cluster boundaries (CMFs) and other data is believed to be more generally useful and more classification-efficient than the operational families. Because such a division also would be quite compatible to the current system, it would ease the transition to the new system being proposed for operational use.

Previous research also showed that there was a small, acceptable reduction in MPP (not statistically significant) when families formed by the CE empirical clustering algorithm (H_d) were modified by judgment as number of families (m) increased from 9 to 25, the reduction being smaller as m is increased. This finding increased our confidence in the use of judgment to attach small entry-level MOS, without validity data, to MOS in the first tier. These small MOS, as noted earlier, would utilize the same AVs used for the family to which it was attached.

For the second tier of the present investigation, we compared 10, 13 and 17 job families clustered by constrained H_d and adjusted by judgment. This included the shredding of the current 10 families to 13 and 17 families. The unbiased, overall MPP of .186 for the 17 families was found to be superior to the 10 or 13 families with MPPs of .149 and .168, respectively.

Quality Constraints

The major constraint imposed on the optimal allocation model for both tiers was the use of job quotas. The proportion of enlistees allocated to jobs or job families in the simulations was proportional to

the actual accession numbers used operationally. The adoption of this realistic constraint reduced MPP values since validities are higher for technical jobs, with smaller quotas, compared to validities that are lower for combat MOS, with larger quotas.

Another major second-tier constraint placed on the optimal allocation process was intended to prevent MPPs for any job family dropping below a minimum MPP value of zero (for statistical standard scores with a mean of zero in the Army input population). This constraint resulted in a considerable reduction of the overall MPP from .145 to .106. Other types of constraints that might result in less loss of MPP were suggested in the previous section. Whatever constraint that is considered desirable for operational use could readily be built into the allocation process. It should be noted that reduction in MPP in the second tier would not reduce the CE of the job assignment process since classification is accomplished in the first tier. Again, the purpose of the second-tier system is to assist recruiters in counseling enlistees, to establish minimum cut scores for jobs, and for record-keeping purposes.

Another constraint employed in the second tier was the use of only positive weights in the computation of LSEs for each family. Since the second tier is visible, we felt that the use of negative weights might be poorly understood by users. Employing only positive weights, however, reduces MPP values by only a trivial amount, .142 vs .145 for the 17-family set.

Summary of Findings Critical to Recommendations

Other highly relevant findings of the present research supporting the recommendations include: (1) the consistency of findings across different investigations using LSEs as AVs; (2) the lack of impact on MPP using groups varying in experience during the first tour of duty; (3) the rapid loss of MPP in unbiased estimates compared to biased estimates using H_d as m exceeds 40; (4) the relatively small gain in unbiased MPP after 40 families using H_d; (5) the effectiveness of judgment in modifying the H_d empirical clusters; (6) the effectiveness of the H_d algorithm in forming empirical families even using MOS with smaller validity samples than previously considered appropriate compared to relying on only MOS with very large sample sizes; and (7) the inability to find a feasible solution to eliminate negative family MPPs (values of family mean AAs below the mean of all AAs for the total Army input population) when using operational AAs.

Highly relevant findings of previous investigations supporting the recommendations include: (1) an acceptable degree of criterion equivalence exists between SQT and CTP, resulting in the same decisions being made using either criterion; (2) unbiased designs are especially important in eliminating the effect of sampling error on MPP as m increases; (3) regression weights applied to predicted standard scores to form LSEs provide higher MPPs than unit-weighted test composites even if the predictors are optimally selected

for each composite; and (4) the mean intercorrelation, r, among AVs has more impact on MPP than does the mean validity, R, of the same AVs, indicating why the best three-test, unit-weighted composites provide higher MPP than the best four-test unit-weighted composites; (5) although there is a distinct underprediction of performance for females and blacks, these differences are so small as to be considered of little practical importance, especially compared to the differences found for operational AAs; (6) MPP increases as m increases for AVs, without evidence of asymptoting as m exceeds 16 and continuing past 25 to 66; and (7) the relationship between MPP and m holds as well for families based on judgment clustering as for families based on CE empirical clustering.

Transforming ASVAB Scores

As noted in the findings section, the standard score scale proposed for the first tier would provide the same estimate of benefits for the operational system as is found in the experimental simulations by transforming operational ASVAB test scores for each enlistee to statistical standard scores. This is accomplished by using a set of test weights and a constant separately for each AV.

However, since the computation of predicted performance scores for each enlistee is transparent (accomplished in the black box) and the only first-tier operation or product seen by the recruiting counselor is a list of recommended jobs, the use of modified statistical standard scores is preferable. It provides the highest value of MPP obtainable and is simple to implement. It requires only the use of the transformation tables given in Appendix D3. We envision a system where the recruiting counselor will be provided a computer-based printout of a list of the top five jobs for consideration by the recruit. The jobs being recommended to the recruit take into account predicted performance scores and Army operational considerations such as job quotas, quality distributions and starting dates for training programs. The counselor, using this list, negotiates with the recruit. He may show the recruit the scores that he/she obtained for each job cluster in the second tier system (described below). If the recruit's preferences and the counselor's proposals based on the list of top jobs can't be reconciled, a second computer-generated list of the next best five jobs is then discussed, and so on, until there is agreement between the two. Research experience indicates that nearly all recruits accept the counselor's job recommendations when recruits are provided an understanding of the jobs being proposed at the same time that they see their AA scores. Only very infrequently do we envision a situation where a recruit continues to express a strong preference for a job(s) when he/she is judged to be unsuitable for that job on the basis of AV scores.

As noted earlier, the Army conventional standard score (ACSS) scale with an expected mean of 100 and a standard deviation of 20 uniformly computed for all AVs is proposed for the second-tier system. These

70

AV mean values are the same as the existing mean AAs in the youth population, before selection into the Army. The use of the ACSS scale necessitates a change in the computation of AVs as compared to AAs because the unit-weighted composites of four ASVAB tests are being replaced by LSEs composed of all ASVAB tests. A given AV score using the ACSS scale in the proposed new system would have the same meaning and use as an AA score in the current operational system. The values for transforming operational test scores into statistical standard scores are given in Appendix E3.

It should be noted again that in using the same expected mean of 100 and standard deviation of 20 across all AVs for the ACSS scale, there is a considerable reduction of MPP (from .142 to .097) because of the loss of validity based hierarchical effects, the contribution to MPP made by capitalizing on variations in standard deviations proportional to validities. Since only the first tier is used in making initial job assignments based on predicted performance, no more than an acceptable reduction in classification efficiency should result from the use of ACSS in the second tier. The second tier, the reader may recall, is used for counseling recruits and the provision of an AV record accessible to the soldier for establishing cut scores. To the extent that this record is used to guide the soldier and counselors as to where he/she is most qualified, the use of ACSS will reduce MPP by a small amount. The use of the ACSS scale for establishing cut scores for all MOS in a job family would result in establishing less precise cut scores than would be obtained by using modified statistical standard scores. Apart from the effort of differences in scale, the pass/fail record of soldiers will differ depending on whether a cut score is being applied to a first or second tier AV score. But precision in establishing cut scores has only a minimal operational impact since in practice very low cut scores are in current use permitting the vast majority of recruits to qualify for most jobs. Additionally, the new system being proposed assigns recruits to jobs for which individuals have their highest predicted performance scores and thus minimum requirements rarely affect classification decisions associated with initial assignments.

Optimal Assignment of Enlistees to Jobs

The unconstrained optimal assignment of recruits to jobs is obtained by assigning individuals to the jobs for which they have their highest predicted performance (PP) scores. Unfortunately, constraints need to be imposed so that operational requirements and policies are met. Any constraint on the optimal assignment algorithm, of course, effects MPP values. For example, when quality distribution assignments to jobs alter assignments, the MPP is decreased. Brogden (1946) proposed a means of optimally assigning individuals to meet job quotas and maximize total MPP by assigning each individual to his/her highest adjusted PP score. Adjusted PP scores are defined as PP scores after the target PP score has been subtracted plus an added job

constant (or dual parameter). In general, jobs with larger quotas and jobs whose PP scores have the highest average correlation with other PP scores tend to have the largest dual parameter.

Making assignment decisions for a recruit without waiting to accommodate a large enough pool to use a batch algorithm is sometimes referred to as a person-by-person algorithm or line-by-line assignment or as sequential assignment. A number of practical person-by-person assignment algorithms exist including: Brogden's (1946b, 1954b) concept of additive column constants; Horst's (1960) and Sorenson's (1965b) use of a multiplier or transformation matrix based on covariances among test composites and validities; Ward's (1958) disposition index; and Cardinet's (1955) graphical profile display. The multidimensional scaling system (MSS) has also been described by the authors as a means of accomplishing person-by-person assignment.

There are more recent and ongoing operational research technologies that might be applied to implement a two-tiered system. For example, a technique employed in the present investigation imposed a quality distribution constraint on optimal assignment where the objective function was MPP. The EPAS system now under development by ARI could further explore programs that force an accommodation between Army constraints, optimization of predicted performance, recruiting estimates, and population constraints and inputs.

Value of Performance Gains

In the present research we found that the selection and classification processes together contribute a total MPP value of .362. The independent contribution of selection is 44% and the contribution of classification is 56%. The total process provides more than twice as much gain in predicted performance as the gain from selection alone. This is a clear indication of the desirability of improving classification efficiency over and above what is now available in ASVAB. Most, if not nearly all, research efforts since the 1970s have focused on improving selection efficiency. Yet earlier research efforts, for example, show that technical and avocational information tests improve CE significantly. Scholarios, et al. (1994) showed that experimental interest and personality measures contribute significant classification efficiency to an expanded ASVAB.

The present research also shows that the first-tier system has an overall MPP more than eight times greater than the existing AA composites (.195 vs .023). Further, the existing AA composites are found to have unacceptable negative MPPs for three of the ten families compared to no unacceptable negatives for the constrained 17 family solution of the second tier.

Even small observed increments in MPP standard scores have been shown to translate into significant and practical estimates of monetary gain. Nord and Schmitz (1991) evaluated the economic benefits of improved MPP across alternative Army assignment strategies. A full least squares model of optimal assignment, which selected and classified individuals with a view to maximizing their predicted performance showed an improvement of .12 in MPP over the operational system currently used by the Army. Using both a net present value (NPV) model of performance valuation (Brogden, 1951; Schmidt & Hunter, 1983) and a more traditional opportunity cost model, Nord and Schmitz estimated the net economic value and cost of this gain. The opportunity cost approach was used because of the subjectivity and possible unreliability of NPV and because of concern when NPV procedures are applied to public sector activities where no clear valuation of output is possible. Under both models, MPP standard score gains of less than .1 were shown to imply significant and practical gains from an improved system of selection and classification. Nord and Schmitz reported dollar gains in the hundreds of millions for the opportunity cost model.

Recommendations

Two major recommendations are made that, if accepted, would have profound implications for the Army's current operational enlisted classification system:

1. Replace the current operational system with the proposed two-tiered system. Use the 150 job families for the first tier and the 17 job families for the second tier.

2. Managers of the personnel selection and classification system need to be made aware of the potential gains in productivity of the proposed system.

The Two-Tiered System

The proposal calls for the establishment of a "black-box" first-tier system in which separate full least squares assignment composites are computed for 150 job families formed from 170 MOS. The first tier is used to propose initial job assignments for presentation to recruits by recruiting counselors. Predicted performance as used within the black-box is transparent. The output of the black-box is a computer-generated list of "best" job suggestions for presentation to the recruit based on predicted performance and other factors such as job quotas, quality distributions and policy constraints. Test AA scores obtained for the second-tier system may be used by the counselor in negotiating an assignment with the recruit.

The set of 17 families of the second tier is used for counseling, establishing cut scores and administration and is recorded on each enlistee's personnel record. The ACSS scale closely resembles the existing AA system with its mean of 100 and SD of 20 and also has essentially the same meaning, justifying the use of existing minimum cut scores intended for use with AAs. The AA areas are associated with families that are familiar because of their use in the operational context, except that the 17 families being proposed are more rational and more homogeneous. The significant difference in the second tier system is that it has much more classification efficiency than the operational AAs as measured by MPP (.097 vs .023). Thus, the 17 family system more accurately performs its intended functions.

The report also provides AVs for all the remaining 80 or so jobs not used in the analysis because of lack of data. These small-sized, entry-level jobs were linked to MOS in the first tier on the basis of judgment and data from other investigations. The same linkages of these 80 jobs in the first tier are also used in expanding the second tier, thus consistently assigning all MOS to both first and second tiers.

The potential MPP for the first-tier system is more than eight times greater than the existing AA composites (.195 vs .023). Initially, this very large magnitude of gain may not appear credible until an

74

account of the development of the operational AA system is understood. The current operational Army enlisted classification and person-job matching systems utilize a set of nine aptitude area test composites corresponding to nine job families (excluding the very small GT aptitude area). This type of system evolved in all the military services after three decades of research emphasis on enhancing predictive validity. The content of both test composites and the ASVAB was selected to maximize predictive validity with little or no attention given to improving the classification efficiency of the total set of test composites in a multiple-job, optimal assignment situation. Traditionally, the number of tests in each composite was kept small and the weights restricted to unity in order to simplify the operational computation and use of the composites. The emphasis on predictive validity and the belief in its operational superiority, in a precomputer age, is shown to be outdated and fundamentally erroneous with respect to both empirical results and psychometric theory.

The authors, in a series of basic and applied research efforts, including the present research, demonstrate that considerable classification efficiency is potentially obtainable from the existing ASVAB if it is used in accordance with differential assignment theory principles. The gain in MPP over the existing AA in the first tier is attributable to (1) the use of full-least squares composites (using all nine best-weighted tests of ASVAB; (2) the use of job families that are appropriately structured (smaller differences among LSEs for jobs within families and larger differences between LSEs for jobs across families; and (3) the non-trivial multidimensionality of the joint predictor-criterion space, an attribution related to the FLS equation in (1) above.

Stability of Regression Equations

The benefits obtainable from the use of full regression equations (FLS), where the benefits are measured in terms of MPP, has received very little attention until recently. Scholarios et al. (1994) provides evidence in a model sampling experiment based on Army Project A data. Assignment variables consisting of regression equations based on ten "best" predictors are clearly superior to such equations based on five "best" predictors, and the selection of predictors using classification-efficient rather than selection-efficient indices provides a better battery for classification.

These results are obtained because the basis of superiority is determined in terms of MPP rather than predictive validity. Most research and professional practice is often guided by predictive validity. Individuals experienced in such practice are often persuaded that the effect of additional variables, or of regression weights instead of unit weights, on predictive validity also applies to classification efficiency.

The rigorous experimental design controlling sampling error employed by Johnson et al. (in preparation) and in the present investigation provide unbiased estimates of MPP for the two-tiered systems.

75

They compare the operational unit-weighted composites with those resulting from FLS composites. These comparisons show the clear superiority of FLS composites. The use of these large independent samples allow for sampling error to take its full toll of shrinkage across the three samples and still the unbiased estimates show that the weights were sufficiently stable to produce impressive gains over the unit-weighted composites.

Job Family Structures

Increasing the number of assignment composites and associated job families adds to potential classification efficiency. Often benefits are obtainable from restructuring the Army job-family structure for classification. In an earlier investigation, Johnson, Zeidner and Leaman (1992) provide strong evidence that clustering of MOS to form new job families that maximized H_d would improve the classification efficiency of the system. The investigation also indicates that CE would be improved by increasing the number of families. In demonstrating the value of a two-tiered system, the number of job families is expanded to 66 families. Again results show an increase in CE with very large increases in the number of jobs. In the present research, where a separate LSE is used for a large number of MOS, an ideal situation for CE, we again find a large increase in CE resulting from expanding the number of families. We, however, expect a much smaller increase in MPP when the number of job families are increased greatly compared to a modest increase in the number of job families. As job families are added to the family structure, there is an appreciable increase in the dimensionality of the joint predictor-criterion space. But as the number of job families expand beyond 50 or 60, there is apparently a relatively smaller increase in the joint space, resulting in a smaller increase in MPP.

Multidimensionality of the Joint-Predictor Criterion Space

DAT is compatible with the prevailing evidence that a single principal component factor explains 75% to 85% of the total factor contribution in the joint predictor-criterion space. The remaining percentage can provide the basis of non-trivial classification efficiency (Statman, 1992). The pool of experimental predictors utilized by Sorenson (1965a) and Scholarios et al. (1994) also supports the significance of multidimensionality in the joint space. In the present research the nine test composites of ASVAB, appropriately weighted, are compared to the operational unit-weighted, four-test composites, and the FLS composites with the larger number of tests capitalizing on multidimensionality are much superior to the AA composites.

The present investigation demonstrates the practical value of regression weights to form AVs, the useful multi-dimensionality of the ASVAB in the joint predictor-criterion space and the value of using full LSEs as AVs. The rigorous experimental design using three independent samples provided unbiased

estimates of MPP. Predicted performance (AV) scores after shrinkage across three large samples still provide sufficiently stable weights to produce impressive MPP gains. The research also demonstrates the benefits of using full (nine test) regression equations when benefits are measured in terms of MPP, in contrast to the lack of further contributions to predictive validity after the selection of the best three or four tests. The gain is also attributable to the use of better structured and more numerous job families.

Potential Benefits of the Proposed System

The operational ASVAB AA composites are shown in the present investigation and prior investigations to be of only minimal value and, further, the existing AA composites are found to have several unacceptable negative MPPs. We have not been able to find a feasible constraint to impose on the optimal allocation algorithm to remove these negatives when AA composites and operational families are used. In contrast, the proposed system has an MPP of .195 compared to .023 for the current system and constraints can be imposed to remove undesirable negatives with some loss in overall MPP. It is important to emphasize that the proposed classification system provides more than twice as much gain in predicted performance as the gain from selection alone. While the selection system supported by the ASVAB is widely extolled by users and researchers, the potential gains of classification appear to be largely unknown to them. Yet, the same gain (or more) from classification is readily available in ASVAB and potentially brings about important gains in productivity with little risk and with little implementation costs.

Cascio (1991) writes that much of what researchers do in the human resources field is largely misunderstood and underestimated by the organizations that researchers serve because "...much of what we do is evaluated only in statistical or behavioral terms. Like it or not, *the language of business is dollars, not correlation coefficients.*" Nord and Schmitz (1991) using a conventional opportunity cost model estimated that gains as small as .1 in MPP resulted in very significant dollar gains. Policy makers and managers of the personnel selection and classification system need to be aware of the dollar gains in productivity and they also need to be aware of the limitations of the present operational system.

There are several potential difficulties in implementing the proposed system including: (1) the reluctance on the part of management to make changes, especially since the proposed system represents the first major change in AA composites since 1949; (2) the proposed system requires the use of an operational EPAS system to implement the two-tiered system in a manner that will maximize MPP and meet changing operational constraints and policies that occur throughout the year; (3) an infrastructure needs to be developed to support the proposed system; (4) the acceptance of the new system requires a good understanding of cost-benefit analysis or utility; and (5) counseling procedures for the proposed system need

to be developed in a manner that insures that potential recruits accept the new system as an aid in job matching, not as a barrier to being assigned to preferred jobs.

Three major technical recommendations are made:

1.   Modify the proposed system to accommodate changes in jobs and in the Army input population with the passage of time.

2.   Improve the race and gender fairness of the ASVAB and its test composites.

3.   Improve the classification effectiveness of the ASVAB by identifying classification-efficient measures and by developing a more CE operational battery.

Managing the System

In the course of time, say within a decade, the technical content of tasks within a job or MOS will change. Because of changes in content, MOS may be dropped from the MOS structure or added to it, some MOS may be merged with other MOS and still other MOS may be partially or radically changed. The Army input population, too, may be changing. Recruits may be more diverse in terms of race, gender, or national origin. They may be less or more able depending, in part, on the attractiveness of the Army to potential recruits compared to other available job and educational opportunities.

It is clear, then, that the validities of ASVAB tests need to be checked for changes in job content or job family structure and for changes that may be brought about because of demographic characteristics. Since the first-tier system consists almost entirely of stand-alone MOS, it is relatively easy to change the AVs of specific MOS based on validity data. To keep the system current a means of collecting and analyzing data is required for maintenance purposes. The system must use technically credible criterion measures. Without a maintenance system, some decay in predictive effectiveness should be expected. ARI needs to find the resources not only to implement the proposed new classification system but also to monitor the present system on a routine basis. In sum, the proposed system needs to accomodate changes in the job structure, characteristics of recruit quality, and incentives, technological and doctrinal innovations.

Race and Gender Fairness

In a previous investigation, Zeidner et al. (1998b) provide evidence that FLS composites have at least as much fairness for black and female recruits as does the current operational system. Fairness is traditionally defined as the absence of underpredictions for minority groups in which discrimination potentially exists. Comparisons of the existing AA composites with the FLS composites for use in classification result in much smaller prediction errors for the latter in all groups.

Nonetheless, there is a distinct pattern of underpredictions for females and blacks. While these differences are so small in most MOS as to be considered of little practical importance, efforts should be made to increase fairness to minorities. Underprediction in gender fairness appears to be better understood than in race fairness and promising techniques, including the use of moderator variables, may assist in improving fairness. Such research appears very important as public policy and technically worthwhile to undertake. Gender fairness research may also lead to developing effective approaches for improving race fairness as well.

Classification-Efficient Tests

Classification-efficient tests should have their highest corrected validities in the job family or the job for which the tests are developed and an appreciably lower average validity in all other job families. A heuristic can be used to form a continuum that has, at one end, high validity generalization and, at the other end, high classification efficiency. Research at ARI in the 1960's focusing on experimental technical and avocational information tests showed that such measures have great promise in producing differential validity. However, in more recent times, when predictive validity was used as the measure of effectiveness, information tests began to disappear from experimental test pools. Research now needs to be directed at identifying classification-efficient tests that are also gender and race fair. No known past research has had this dual objective.

References

Abbe, E. N. (1968). *Statistical properties of allocation averages* (Memorandum 68-13). Washington, DC: U.S. Army Behavioral Science Research Laboratory.

Boldt, R. F., and Johnson C.D. (March 1963). *Computerized Personnel Assignment System.* Paper presented at the Second Army Operations Research Symposium, Durham, North Carolina.

Brogden, H. E. (1946). An approach to the problem of differential prediction. *Psychometrika, 11,* 139-154.

Brogden, H. E. (1954). A simple proof of a personnel classification theorem. *Psychometrika, 19,* 205-208.

Brogden, H. E. (1955). Least squares estimates and optimal classification. *Psychometrika, 20,* 249-252.

Brogden, H. E. (1959). Efficiency of classification as a function of number of jobs, percent rejected, and the validity and intercorrelation of job performance estimates. *Educational and Psychological Measurement, 19,* 181-190.

Cardinet, Jean (1959) The use of profiles for differential classification. *Educational and Psychological Measurement, 19,*191-205.

Dantzig, G. B. (1951a). Maximization of a Linear Function of Variables Subject to Linear Inequalities. In T.C. Koopmans (Ed.), *Activity Analysis of Production and allocation* (chapter 21). New York: Cowles Commission Monograph 13, John Wiley.

Dantzig, G. B. (1951b) Application of the Simplex Problem to a Transportation Problem. In T.C. Koopmans (Ed.), *Activity Analysis of Production and allocation* (chapter 23). New York: Cowles Commission Monograph 13, John Wiley and Sons, Inc.

Dwyer, P. S. (1954). Solution of the personnel classification problem with the method of optimal regions. *Psychometrika, 19,* 11-26.

Dwyer, P. S. (1957). The detailed method of optimal regions. *Psychometrika, 22,* 43-52.

Ford. L. R., & Fulkerson, D. R. (1962). *Flows in Networks.* Princeton: Princeton University Press.

Gass S. I. (1975). *Linear programming: Methods and applications.* New York: McGraw-Hill.

Granda, T. M., & McMullen, R. L. (1974, October). *A heuristic approach to a large personnel assignment problem.* Paper presentation at the Fall ORSA/TIMS National Convention, San Juan, Puerto Rico.

80

Hatch, R. S. (1966). *ACT II system: System memorandum.* Rockville, MD: Decision Systems Associates Inc.

Hatch, R. S. (1969). Development of optimal allocation algorithms for personnel assignment. *Proceedings of NATO conference, Mathematical Models for the Management of Manpower Systems, Porta, Portugal,* London: English Universities Press.

Hatch, R. S. (1971). Development of optimal allocation algorithms for personnel assignment. In A.R. Smith (Ed.), *Models of Manpower Systems* (pp. 383-397). New York: Elsevier Publishing Co., Inc.

Horst, P. (1954). A technique for the development of a differential predictor battery. *Psychometrika, 68* (9, Whole No. 380).

Horst, P., & Sorenson, R. C.(1976, December). *Matrix transformations for optimal personnel assignments* (TN 77-5). San Diego: Navy Personnel Research and Development Center.

Johnson C. D., & Sorenson R. C. (1974). Model sampling experimentation for manpower planning. In D. J. Clough, C. J.Lewis, & A. L. Oliver (Eds.), *Proceedings of a NATO conference held in Cambridge UK, Manpower Planning Models* (pp. 37-52). London: English University Press.

Johnson, C. D., & Zeidner, J. (1991). *The economic benefits of predicting job performance. Vol. 2: Classification efficiency.* New York: Praeger.

Johnson, C. D., & Zeidner, J. (1995). *Differential assignment theory sourcebook* (Research Note 95-43). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Johnson, C. D., Zeidner, J., & Leaman, J. A. (1992). *Improving classification efficiency by restructuring Army job families* (Technical Report 947). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Johnson C. D., Zeidner, J., Scholarios, D. (1996). *Selection for multiple jobs from a common applicant pool.* (Draft Research Report). Alexandria, VA: The Army Research Institute.

Johnson, C. D., Zeidner, J., & Vladimirsky, Y. (in preparation). *Developing classification-efficient job families using differential assignment theory techniques.* (Research Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Koopmans, T. C. (Ed.). (1951). *Activity analysis of production and allocation.* New York: Cowles Commission Monograph 13, John Wiley.

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly, 2,* 83-97.

Kuhn, H. W. (1956). Variants of the Hungarian method for assignment problems. *Naval Research Logistics Quarterly, 3,* 253-258.

Nord, R., & Schmitz, E. (1991). Estimating performance and utility effects of alternative selection and classification policies. In J. Zeidner & C.D. Johnson. *The economic benefits of predicting job performance. Vol. 3: The gains of alternative policies.* New York: Praeger.

Scholarios, D., Johnson, C. D., & Zeidner, J. (1994). Selecting predictors for maximizing the classification efficiency of a battery. *Journal of Applied Psychology, 79,* 412-424.

Sorenson, R. C. (1965, August). *Development and evaluation of a matrix transformation useful in personnel classification.* Unpublished doctoral dissertation, University of Washington, Seattle.

Statman, M. A. (1992, August). *Developing optimal predictor equations for differential job assignment and vocational counseling.* Paper presented at the American Psychological Association Annual Meeting, Washington, DC.

Stichman, E. P. (1963, April). *A linear program for transportation cost minimization of BCT-AIT Shipping.* Washington, DC: U.S. Army Personnel Research Office.

Ward, J. H., Jr. (1958). The counseling assignment problem. *Psychometrika, 23,* 55-56.

Ward, J. H., Jr. (1959, April). *Use of a decision index in assigning Air Force personnel* (WADC-TN-59-38). San Antonio: Air Force Human Resources Laboratory.

Whetzel, D. L. (1991). *Multidimensional screening: Comparison of a single-stage personnel selection/classification process with alternative strategies.* Unpublished doctoral dissertation, The George Washington University, Washington, DC.

Zeidner, J., & Johnson, C. D. (1991a). *The economic benefits of predicting job performance. Vol.1: Selection utility.* New York: Praeger.

Zeidner, J., & Johnson, C.D. (1991b). *The economic benefits of predicting job performance. Vol.3: Estimating the gains of alternative policies.* New York: Praeger.

Zeidner, J., & Johnson, C.D. (1991c). Classification efficiency and systems design. *Journal of the Washington Academy of Sciences, 81,* 110-128.

Zeidner, J., & Johnson, C.D. (1994). Is personnel classification a concept whose time has passed? In M.G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification: New directions.* Hillsdale, NJ: Erlbaum.

Zeidner, J., Scholarios, D, & Johnson, C.D. (1997). Evaluating military selection and classification systems in the multiple job context. *Military Psychology, 9,* 169-186.

Zeidner, J., Johnson, C.D., & Vladimirsky, Y. (1997). *The substitutability of criteria in the development and evaluation of ASVAB classification procedures.* (Technical Report 1071). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Zeidner, J., Johnson, C.D., & Vladimirsky, Y. (1998b, in preparation). *Fairness of proposed new Army ASVAB test composites for restructured job families.* (Research Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.